18 May 2012 | $10

Science

AAAS

SPECIAL SECTION

Human Conflict

page 796

CONTENTS continued >>

COVER

We are all too familiar with human conflict, such as this bombing on 13 August 2006 in Beirut, Lebanon, by the Israel Defense Forces. In the special section on Human Conflict (see page 818), we examine the origins of conflict, trace its path through history, and consider its modern manifestations. We also analyze our innate ability to foster peace and look at societies that eschew war. This cover was chosen for visual imagery and not for any political message or endorsement.

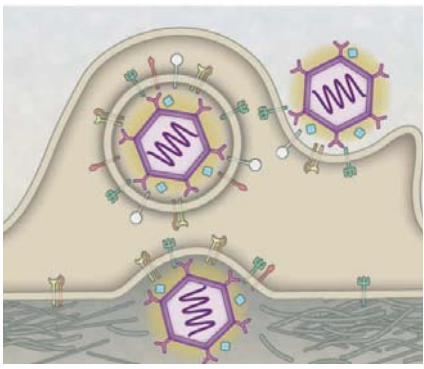
Image: Sang-Hoon KISH Kim/Sipa Press/Newscom

DEPARTMENTS

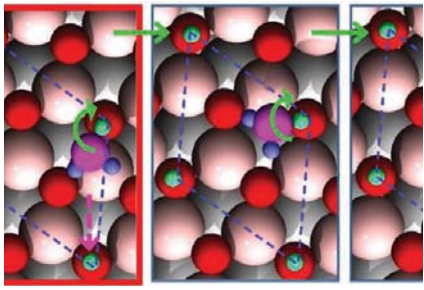
PERSPECTIVES

BREVIA

REPORTS



page 809



page 889

CONTENTS continued >>

SCIENCEONLINE

SCIENCEXPRESS

www.sciencexpress.org

Evolution and Functional Impact of Rare Coding Variation from Deep Sequencing of Human Exomes
J. A. Tennessen et al.
Most functionally consequential variants in protein-coding genes are rare and, thus, difficult to find.
10.1126/science.1219240

An Abundance of Rare Functional Variants in 202 Drug Target Genes Sequenced in 14,002 People
M. R. Nelson et al.
A pharmacogenomics analysis shows how challenging it will be to associate rare variants with phenotypes.
10.1126/science.1217876

Oscillatory Dynamics of Cdc42 GTPase in the Control of Polarized Growth
M. Das et al.
The regulation of a yeast cell-growth enzyme is dynamic rather than on-off.
10.1126/science.1218377

Graphene Barristor, a Triode Device with a Gate-Controlled Schottky Barrier
H. Yang et al.
The absence of defects and surface oxides at a graphene/silicon interface enables voltage control of graphene devices.
10.1126/science.1220527

Direct Detection of Projectile Relics from the End of the Lunar Basin–Forming Epoch
K. H. Joy et al.
Analysis of lunar rocks from the Apollo missions reveals fragments from meteorites that hit the Moon in the ancient past.
10.1126/science.1219633

Fragments of the Lunar Cataclysm
A. E. Rubin
10.1126/science.1224184

Roton-Type Mode Softening in a Quantum Gas with Cavity-Mediated Long-Range Interactions
R. Mottl et al.
Low-energy excitations of the type present in superfluid helium are observed in a cold gas of rubidium atoms.
10.1126/science.1220314

SCIENCENOW

www.sciencenow.org
Highlights From Our Daily News Coverage

Electronics Go Viral
Researchers create an electrical generator out of viruses.
http://scim.ag/Electronics_Viral

Engravings of Female Genitalia May Be World's Oldest Cave Art
Images found on the ceiling of a collapsed shelter may predate those of France's famed Chauvet Cave.
http://scim.ag/Cave-Art

Heat Trickery Paves Way for Thermal Computers
Devices harness heat currents instead of electrical ones.
http://scim.ag/Thermal-Computers

SCIENCESIGNALING

www.sciencesignaling.org
The Signal Transduction Knowledge Environment
15 May issue: http://scim.ag/ss051512

RESEARCH RESOURCE: MicroSCALE Screening Reveals Genetic Modifiers of Therapeutic Response in Melanoma
K. C. Wood et al.
Miniaturized functional genomics uncovers potentially useful combination therapies and drug targets for melanoma.

PERSPECTIVE: The Structure of the Full-Length Tetrameric PKA Regulatory Riβ Complex Reveals the Mechanism of Allosteric PKA Activation
J. M. Elkins and S. Knapp
The conformational changes caused by cAMP binding to the regulatory subunit in the PKA holoenzyme are shown.

PERSPECTIVE: The Structural Basis of DKK-Mediated Inhibition of Wnt/LRP Signaling
J. Bao et al.
Structural analysis indicates that the binding sites on a receptor for ligand and inhibitor partially overlap.

PRESENTATION: Roles of GRK2 in Cell Signaling Beyond GPCR Desensitization—GRK2-HDAC6 Interaction Modulates Cell Spreading and Motility
P. Penela et al.
GRK2 modulates tubulin acetylation dynamics in an HDAC6-dependent manner to affect epithelial cell spreading and motility.

SCIENCETRANSLATIONAL MEDICINE

www.sciencetranslationalmedicine.org
Integrating Medicine and Science
16 May issue: http://scim.ag/stm051612

EDITORIAL: APOε4 Status and Traumatic Brain Injury on the Gridiron or the Battlefield
S. Gandy and S. T. DeKosky
Should APOε4 genotyping be required for high school or college athletes and military personnel?

RESEARCH ARTICLE: Chronic Traumatic Encephalopathy in Blast-Exposed Military Veterans and a Blast Neurotrauma Mouse Model
L. E. Goldstein et al.
Blast exposure is associated with chronic traumatic encephalopathy and persistent cognitive deficits in veterans and athletes.
>> News story p. 790

RESEARCH ARTICLE: Gene Therapy for Aromatic L-Amino Acid Decarboxylase Deficiency
W.-L. Hwu et al.
Gene therapy can restore some motor function in patients with aromatic L-amino acid decarboxylase deficiency.

RESEARCH ARTICLE: CD25 Blockade Depletes and Selectively Reprograms Regulatory T Cells in Concert with Immunotherapy in Cancer Patients
A. J. Rech et al.
CD25 monoclonal antibody therapy rapidly and durably depletes T_{regs} in cancer patients through a mechanism consistent with reprogramming.

RESEARCH ARTICLE: High-Throughput Sequencing Detects Minimal Residual Disease in Acute T Lymphoblastic Leukemia
D. Wu et al.
High-throughput sequencing can detect minimal residual disease comparable to multiparametric flow cytometry in T-ALL patients.

PERSPECTIVE: Drug-Based Optical Agents—Infiltrating Clinics at Lower Risk
W. Scheuer et al.
Using drugs as optical imaging agents and "microdosing" amounts reduces risk or clinical translation of fluorescence molecular imaging.

SCIENCECAREERS

www.sciencecareers.org/career_magazine
Free Career Resources for Scientists

Winning Over Hearts Means Understanding Minds
M. Price
For scientists who study conflict's motivations and consequences, the brain is the battlefield that matters.
http://scim.ag/HeartsandMinds
>> Human Conflict section p. 818

Tooling Up: How to Swim With Sharks
D. Jensen
Here's what you need to know about third-party services that offer to help with your job search.
http://scim.ag/TU_Sharks

In Person: Family-Friendly Science Careers
T. Ainsworth
We need to let young women know that it *is* possible to have a science career and a family.
http://scim.ag/InPersonAinsworth

SCIENCEPODCAST

www.sciencemag.org/multimedia/podcast
Free Weekly Show

On the 18 May *Science* Podcast: a special show exploring human conflict, including strife among our primate ancestors, the biological underpinnings of racism, and the fundamentals of "peace systems."

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals Mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Domestic individual membership and subscription (51 issues): $149 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $990; Foreign postage extra: Mexico, Caribbean (surface mail) $55; other countries (air assist delivery) $85. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #1254 88122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A.**

Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $10.00 current issue, $15.00 back issue prepaid includes surface postage; bulk rates on request. **Authorization to photocopy** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act is granted by AAAS to libraries and other users registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that $30.00 per article is paid directly to CCC, 222 Rosewood Drive, Danvers, MA 01923. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

AAAS
ADVANCING SCIENCE, SERVING SOCIETY

EDITORS' CHOICE

EDITED BY KRISTEN MUELLER AND JAKE YESTON

ECOLOGY

Different Dialects

Acoustical analysis has revealed the presence of complex communication signals across a variety of animal species. Information may also be conveyed by the arrangement of sounds, known as syntax. Syntax is thought to be rare outside of humans and birds, but the ability to transmit complex information vocally is likely to be beneficial across social species. Male rock hyraxes, a type of social small mammal, make a complex call that consists of wails, chucks, snorts, squeaks, and tweets and conveys detailed information about the caller's identity. Kershenbaum *et al*. applied methods borrowed from information theory and genetics to show the existence of distinct syntactical dialects in male hyraxes across regions within Israel. Specifically, populations near to one another were more similar than those farther apart, whereas at larger distances, differences varied without regard to proximity. The breakdown of the relationship between dialect and distance at larger scales suggests that dispersal distance is a limiting factor in the transmission of dialect learning. These results demonstrate that syntactical structure and vocal learning may occur in a more diverse group of species than has been previously recognized and that sociality may be a driving force for its development. — SNV

Proc. R. Soc. London Ser. B **279**, 10.1098/rspb.2012.0322 (2012).

PSYCHOLOGY

Evaluating Rituals

Rituals—protocols designed to make a problem go away or to bring about a favorable outcome—are widely used across cultures yet causally inexplicable within the mechanistic schema of the physical world. Legare and Souza collected judgments of ritual effectiveness from patrons of public health centers in Belo Horizonte, Brazil. They found, not surprisingly, that individuals who had themselves used rituals (referred to as simpatias) reported a stronger belief in their effectiveness. They also found, via experimental manipulation of scenario simpatias, that three characteristics contributed significantly to these judgments: (i) the number of steps in the protocol to be performed; (ii) the repetition of one or more steps in the ritual; and (iii) the inclusion of religious icons. Taken together, these findings are consistent with the proposal that rituals activate an intuitive system of causal belief. — GJC

Cognition **124**, 10.1016/j.cognition.2012.03.004 (2012).

CELL BIOLOGY

Who Hid the Cyclin D2?

During development of the cortex of the mammalian brain, radial glia divide asymmetrically to give rise to apical progenitor cells that continue to divide and cells that differentiate into neurons. Thus, the tissue supports the continued formation of neuronal structures while maintaining a pool of progenitors. Tsunekawa *et al*. report on a mechanism that influences

Continued on page 780

CREDIT: DERKARTS

Published by AAAS

Continued from page 778

the fate of the daughter cells. Radial glial cells have long thin apical and basal processes that extend from either end of the cell. mRNA encoding the cell cycle regulator cyclin D2 was preferentially localized and translated in the basal process because of a regulatory sequence in the 3′ untranslated region of the mRNA. The daughter cell that inherited the basal process thus got most of the cyclin D2 and continued to proliferate. The other daughter cell, perhaps because of a prolonged cell cycle, or effects of other sequestered factors, underwent neuronal differentiation. A causal role of cyclin D2 was supported by experiments depleting or overexpressing the protein, which caused the accumulation of proliferating progenitor cells or increased neurogenesis, respectively. — LBR

EMBO J. **31**, 1879 (2012).

DEVELOPMENT

Stressful for the Long Haul

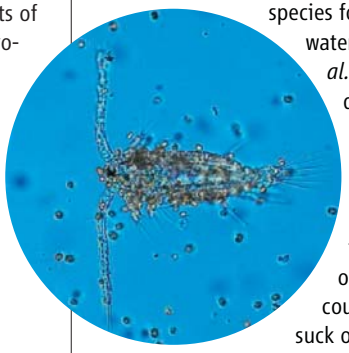
Cold, dehydration, variation in food supply—these are all potential environmental stressors that organisms must face. Often, organisms have mechanisms that can accommodate such challenges, at least to a degree; however, some challenges, particularly those that are rare, can have detrimental effects. Recent studies have reported that developmental changes resulting from such challenges are sometimes passed to subsequent generations through epigenetic mechanisms. Stern *et al*. sought to further examine this using the fruit fly, *Drosophila melanogaster*. *Drosophila* larvae were exposed to toxic stress through supplementation of their food with the drug G418. Stress was applied in a region-specific manner, however, because the flies were engineered to express a resistance gene under the regulation of an arbitrary spatiotemporally restricted developmental promoter. The developmental responses of the flies varied (for example, reduced fly size or deformed wings) as measured by survival rate, larval development, and morphology. Such variation was the result of the resistance genes being under the control of different promoters. The ability to tolerate the stressor was dependent on the down-regulation of Polycomb Group genes. Examination of progeny from toxin-exposed flies, who themselves were not exposed to the toxin, revealed that offspring from several subsequent generations retained the resistance gene before converting back to the normal fly phenotype. These results indicate that stressors may induce both heritable and nonheritable developmental effects. — BAP

Cell Rep. **1**, 10.1016/j.celrep.2012.03.012 (2012).

MICROBIOLOGY

Attack of the Killer Algae

In the oceans, free-living dinoflagellates are one of the most diverse groups of simple eukaryotes: Some are photosynthetic, some are grazers and predators, some are both, and some become temporarily photosynthetic simply by virtue of the cells they eat. This relaxed approach to nutrition is common in the harmful algal bloom species found in coastal waters. But Berge *et al.* have found that dinoflagellate species, which possess permanent chloroplasts and were normally thought to rely on photosynthesis, could attach to and suck out nutriment from crustacean copepod larvae. More usually it is the copepods that are the predators on algae, and their grazing reduces algal biomass and regulates bloom formation, but in this case the dinoflagellate *Karlodinium armiger* displays pronounced swarming behavior and forms feeding aggregates that allow the ingestion of prey much larger than itself. — CA



CREDIT: TERJE BERGE AND NIELS DAUGBJERG

ISME J. **6**, 10.1038/ismej.2012.29 (2012).

PHYSICS

Watching Excitons Condense

A fascinating property of the quantum world is that particles come equipped with a dual, wave-like nature, which becomes apparent at low temperatures and high densities. For a certain type of particle, bosons, this results in the formation of a giant wave, a Bose-Einstein condensate (BEC). BECs have been realized in laser-trapped ultracold gases of alkali elements, but bosons that exist elsewhere in nature are also expected to condense. An exciton is a composite boson consisting of an electron and a hole (a missing electron) in a semiconductor. High *et al.* take a page out of the atomic physicists' book and demonstrate the condensation of excitons in a trap. A spatially varying electric field traps the excitons' dipole moments, which are nonzero because of the spatial separation of electrons and holes in a GaAs/AlGaAs coupled quantum well (such so-called indirect excitons also have the advantage of a long lifetime necessary for cooling). When the temperature is lowered below ~2 K, the excitons collect at the trap center,

Continued on page 782

Published by AAAS

Continued from page 780

and the entire trap volume becomes coherent, signifying the onset of a BEC. The realization of a BEC in a trapped semiconductor system is expected to enable explorations complementary to those possible with ultracold atomic gases. — JS

Nano Lett. **12**, 10.1021/nl300983n (2012).

OCEAN SCIENCE

Where Carbonate Comes From

Rhodoliths—a type of marine red algae that resemble corals—constitute one of the world's most voluminous shallow-water benthic communities. Though they can be found in many tropical locations worldwide, they may be most extensive on the Abrolhos Shelf off the coast of Brazil. Amado-Filho *et al.* conducted a detailed survey of those beds, in order to determine rhodolith distribution, extent, composition, and structure. They found that the beds cover an area of about 20,900 km^2, comparable to that of the Great Barrier Reef, Australia. They estimated that the mean rate of $CaCO_3$ production by these organisms is about 0.025 Gt/year, in line with the world's largest biogenic $CaCO_3$ deposits, and that the rhodolith beds of the Abrolhos Shelf alone contain approximately 5% of the $CaCO_3$ inventory of all the world's carbonate banks. Sedimentation from land-based sources and large-scale dredging and mining are the largest immediate threats to these communities, and ocean acidification presents another, longer-term danger. Rhodoliths accumulate rapidly on a geological time scale, but very slowly on a human time scale, and should thus be considered a nonrenewable resource that needs to be protected. — HJS

PLoS ONE **7**, e35171 (2012).

CHEMISTRY

Tuning the Mix

Metal organic framework materials hook together metal ions or clusters, using organic linkers to make crystalline compounds with two- or three-dimensional porosity. The zeolitic imidazolate framework (ZIF) subset of this class of materials has imidazole molecules as linkers and forms structures analogous to the aluminosilicate lattices of zeolites. In addition to molecular sieving properties, ZIFs can manifest a gating phenomenon whereby interaction with adsorbates leads to an increase in guest uptake. This gating is caused by a rotation of the linker molecules. Thompson *et al.* explored the effects of making ZIFs with a mixture of linking molecules, based on ZIF-8, ZIF-90, and ZIF-7 parent materials, which use 2-methylimidazole, carboxaldehyde-2-imidazole, and benzimidazole linkers, respectively. In particular, ZIF-8 can selectively remove carbon dioxide from a complex mixture of gases, so finding ways to enhance its properties without reducing its stability is of practical importance. Changes in the gate-opening phenomenon emerged for varying specific compositions. Overall, the results indicate that it may be possible to tune the fine structure of a ZIF framework at the synthesis stage, rather than through complicated postsynthetic modifications. — MSL

Chem. Mater. **24**, 10.1021/cm3006953 (2012).

IMMUNOLOGY

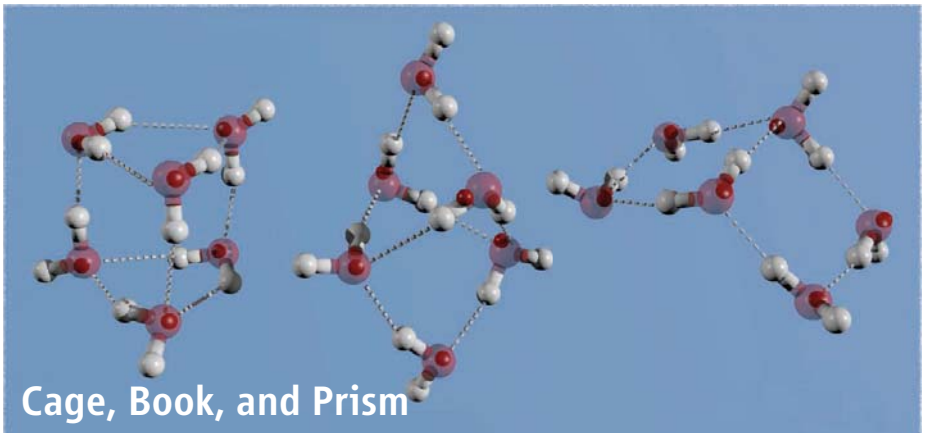
Detecting Danger

Besides responding to infections, the immune system can also recognize tissue injury in the absence of any infectious agents. Examples of this include antitumor immunity and responses to transplanted organs. During these "sterile" responses, the immune system is triggered by so-called DAMPs, danger-associated molecular patterns. These include intracellular contents such as ATP and HMGB1 that are released upon cell damage or death. The C-type lectin DNGR-1 (Clec9a) is a receptor expressed by certain subsets of dendritic cells that is required for the presentation of antigens derived from necrotic cells to T cells. What it recognizes on dying cells, however, has remained a mystery. After a rather challenging hunt, Zhang *et al*. and Ahrens *et al*. now report that DNGR-1, recognized actin filaments. Monomeric actin, or G-actin, was not recognized by DNGR-1, and actin binding proteins such as spectrin enhanced DNGR-1 recognition of actin. The identification of F-actin as a ligand for DNGR-1 reinforces the idea that molecules that normally play a housekeeping role in healthy cells are able to activate the immune system when released into the extracellular milieu. Although this often triggers a controlled immune response that promotes tissue repair, alterations in this response could drive inflammation and contribute to disease processes. — KLM

Immunity **36**, 635; 646 (2012).

Published by AAAS

EDITED BY STELLA HURTLEY



Cage, Book, and Prism

The array of hydrogen bonds governing the extended structure of liquid water is so intricate that chemists have often sought to understand it by studying simpler clusters. Even so, it has been challenging to get a handle on the preferred arrangement adopted by just six water molecules. Interdependent theoretical and spectroscopic studies have narrowed down the lowest-energy hexamer structures to three isomers—respectively designated the cage, the book, and the prism—but their relative energies remain uncertain. Now, **Pérez *et al.*** (p. 897; see the Perspective by **Saykally and Wales**) have observed all three isomers in a single experiment, using Fourier transform microwave spectroscopy, and were able to establish definitively their energy ordering.

Water-Assisted Proton Diffusion

Proton diffusion on metal oxide surfaces can play an important role in many catalytic processes. The presence of water is thought to accelerate proton diffusion. **Merte *et al.*** (p. 889) used high-speed, high-resolution scanning tunneling microscopy to study proton diffusion on an iron oxide. On oxygen-terminated FeO monolayer films formed on Pt, molecular water accelerated proton diffusion. Density function theory calculations implicated a H_3O^+ transition state in the diffusion process.

Mechanisms in Methanol Catalysis

The industrial production of methanol from hydrogen and carbon monoxide depends on the use of copper and zinc oxide nanoparticles on alumina oxide supports. This catalyst is "structure sensitive"; its activity can vary by orders of magnitude, depending on how it is prepared. **Behrens *et al.*** (p. 893, published online 19 April; see the Perspective by **Greeley**) used a combination of bulk and surface-sensitive analysis and imaging methods—along with insights from density functional theory calculations—to study several catalysts, including the one similar to that used industrially. High activity depended on the presence of steps on the copper nanoparticles stabilized by defects such as stacking faults. Partial coverage of the copper nanoparticles with zinc oxide was critical for stabilizing surface intermediates such as HCO and lowering energetic barriers to the methanol product.

Radioactive Resonance

Nuclear magnetic resonance (NMR) spectroscopy and its spatially sensitive cousin, magnetic resonance imaging, have found widespread application in chemical and biological characterization studies. For the most part, these studies take advantage of the energy bifurcation manifested by hydrogen nuclei with oppositely directed spins in a strong magnetic field. More generally, many heavier elements manifest the same effect—including carbon-13, fluorine, and phosphorus. In theory, researchers have known for 50 years that plutonium nuclei have a net spin conducive to NMR. **Yasuoka *et al.*** (p. 901; see the Perspective by **Albrecht-Schmitt**) have now at last observed the resonance of the Pu-239 isotope in a sample of plutonium dioxide.

Keep Your Distance

Conspecific negative density dependence (CNDD), whereby the abundance of a species is limited by negative interactions between individuals of the same species, is thought to have an important influence on the composition and dynamics of forest communities, but studies have generally been limited to few species and small areas. **Johnson *et al.*** (p. 904) analyzed CNDD in over 200,000 plots from a database of more than 3 million individuals of 151 species spanning 4 million square kilometers across forests in the eastern United States and found that the strength of CNDD strongly predicted the relative abundance of tree species. Because tree seedlings are unlikely to become established where conspecific adults are common, CNDD may provide a general mechanism maintaining diversity in forests.

Bring In the Inspectors

In order to assess the impact of occupational and health practices in the state of California, **Levine *et al.*** (p. 907) compared more than 400 uninspected firms with a matched set of inspected firms that were chosen at random. Employees at the inspected firms were less frequently injured and, consequently, the inspected firms suffered fewer injury-related costs. Encouragingly, there were no significant differences in other economic outcomes, such as sales and employment levels, between control and inspected firms.

Ultimate Blockade

A Rydberg atom has an electron in a highly excited energy state, close to being set free, but not quite. Ensembles of such atoms interact strongly, sometimes leading to blockade effects where the excitation of one atom prevents the excitation of another. **Dudin and Kuzmich** (p. 887, published online 19 April; see the Perspective by **Grangier**) demonstrate the generation of a many-body excitation with no more than one Rydberg atom in a mesoscopic ensemble of ultracold atoms. When the principal quantum number was increased beyond 70, the excitation was converted into a photon. The ability to control the creation of excitations provides a promising system for quantum information storage, as well as a source of correlated photons.

Accounting for *Lac*

When *Escherichia coli* expresses the *lac* operon, it needs to balance the potential increase in growth rate conferred through having the encoded proteins (which help it to take up and metabolize lactose) with the potential reduction in growth rate because of the resources needed to express and fold the encoded proteins. To investigate these inherent tradeoffs, **Eames and Kortemme** (p. 911)

Continued on page 773

Published by AAAS

Continued from page 771

investigated the costs and benefits associated with the *lac* operon and surprisingly found that the main constraint on the system is not caused by the manufacture of the proteins themselves but by the lactose permease activity encoded by the *LacY* gene.

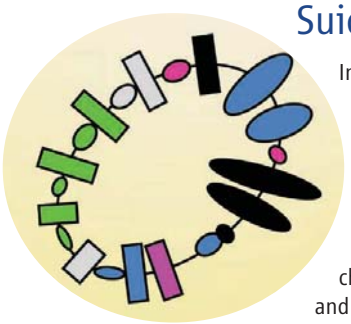
The Hibernating Ribosome

When bacteria enter stationary phase, their ribosomes are inactivated. In *Escherichia coli*, ribosome modulation factor (RMF) causes dimerization of the 70S ribosome and the dimer is stabilized by, hibernation promotion factor (HPF). Alternately, the stationary phase protein, YfiA, inactivates 70S ribosomes. **Polikanov *et al.*** (p. 915) present high-resolution structures of the *Thermus thermophilus* 70S ribosome bound to each of these three factors. The structures suggest that RMF binding inhibits protein synthesis by preventing initial messenger RNA (mRNA) binding and that HPF and YfiA have overlapping binding sites and would both interfere with binding of mRNA, transfer RNA, and initiation factors.

An Aspirin a Day?

The protein kinase AMPK (adenosine monophosphate–activated protein kinase) directly monitors cellular energy stores as reflected by changes in cellular concentrations of AMP, adenosine diphosphate (ADP), and adenosine triphosphate (ATP). Through phosphorylation of its targets, it helps to control metabolism, polarity, autophagy, and the restraint of cell proliferation. Activation of AMPK is also proposed to be beneficial for the treatment of diseases, including cancer and diabetes. **Hawley *et al.*** (p. 918, published online 19 April; see the Perspective by **Shaw and Cantley**) report that AMPK can be activated by high concentrations of salicylate, a compound derived from the very commonly used drug aspirin. In mice, salicylate promoted fatty acid and carbohydrate metabolism in an AMPK-dependent fashion.

Suicidal B cells



In response to an infection, immunological B cells undergo a maturation process that results in the production of immunoglobulin (Ig) that is better able to bind and clear the invading pathogen. This occurs through somatic cell hypermutation and class switch recombination of the Ig gene and requires activation-induced deaminase (AID). **Péron *et al.*** (p. 931, published online 26 April) observed that the 3′ cis-regulatory region of the heavy-chain locus is transcribed and undergoes AID-mediated mutation and recombination. The resulting deletion of the Ig heavy gene cluster generates B cells that are no longer able to express Ig on the cell surface. Because cell-surface Ig expression is essential for B cell survival, this process is termed "locus suicide recombination" (LSR) and may be important in shaping the dynamics of B effector cell differentiation and homeostasis.

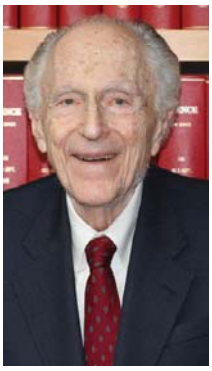
CREDIT: NICOLAS DENIS-LAGACHE, SOPHIE PÉRON, AND MICHEL COGNÉ

Deep Breathing

Living microbes have been discovered many meters into marine sediments. On a cruise in the North Pacific Gyre, **Røy *et al.*** (p. 922) discovered that oxygen occurred for tens of meters into the sediment. The bacteria living in these sediments were respiring the oxygen but at a slower rate than the supply of organic material dropping out of the water column, allowing these ancient deep marine sediments to remain oxygenated. Modeling showed that the rate of respiration of specific carbon decreased as a function of sediment depth, that is, its age. Thus aerobic metabolism can persist in deep marine sediments.

Color and Movement

From humans to insects, color and motion information are thought to be channeled through separate neural pathways for efficient visual processing, but it remains unclear if and how these pathways interact in improving perception of moving colored stimuli. By using sophisticated *Drosophila* genetics, intracellular electrophysiology, two-photon imaging, and behavioral experiments, **Wardill *et al.*** (p. 925) found that early in the processing stage-color photoreceptors influence the motion pathway and that this input improves the flies' optomotor performance in a flight simulator.



David Hamburg is a Visiting Scholar and former president of the American Association for the Advancement of Science and a DeWitt Wallace Distinguished Scholar at Weill Cornell Medical College, New York. E-mail: dhamburg@aaas.org.

Preventing Mass Violence

OUR SPECIES HAS A LONG HISTORY OF DISTRUSTING STRANGERS, DESPISING OUTGROUPS, and fighting each other in many ways and many places, using the most damaging technology available at the time. This special issue of *Science* on Human Conflict illustrates the role of the scientific community in achieving deep insights into the evolutionary history of human conflict and, most importantly, demonstrates how science can help to clarify the major factors that affect the risk of mass violence (see p. 818). Such understandings are critical to develop education that decreases the probability of conflicts, to provide early and active help to countries having trouble with intergroup relations, to help foster equitable democratic development, to protect and promote human rights, and to place constraints on weaponry.

Today, the capacity to incite hatred and violence is greater than ever before. Yet there are promising lines of inquiry and innovation that promote humane, democratic, and prosocial development in childhood and adolescence, as well as responsible adult leadership. It is vital that schools and other child-rearing institutions, including religious organizations, provide a setting in which young people from different backgrounds can overcome ingroup bias and reach valued common goals in an atmosphere of collaboration. Decades of experimental research with youth and adults show that initial animosity can be converted to cooperation if favorable conditions are provided that benefit the groups involved, supported by relevant authorities. There is a strong positive effect of friendly contact when members of two groups work together, with equal status, toward a superordinate goal that can be obtained only through cooperative activity.

A peaceful world will depend on the international community, and especially the established democracies, being on the alert for sister countries in trouble, while being both prepared and willing to extend help to those nations. Many (but not all) countries are open to accepting help in easing intergroup tensions if approached with sympathetic interest and a vision of better opportunities. Integrative solutions can be developed by reformulating basic questions and putting problems into a different context. A high priority can be given to programs of disease prevention that can be accomplished only through the mutual cooperation of groups in conflict. For example, improving the health of children through intergroup cooperation across adversarial boundaries can be used as a point of entry for providing wider help.

Online
sciencemag.org
Podcast interview (http://scim.ag/ed_6083) with author David Hamburg.

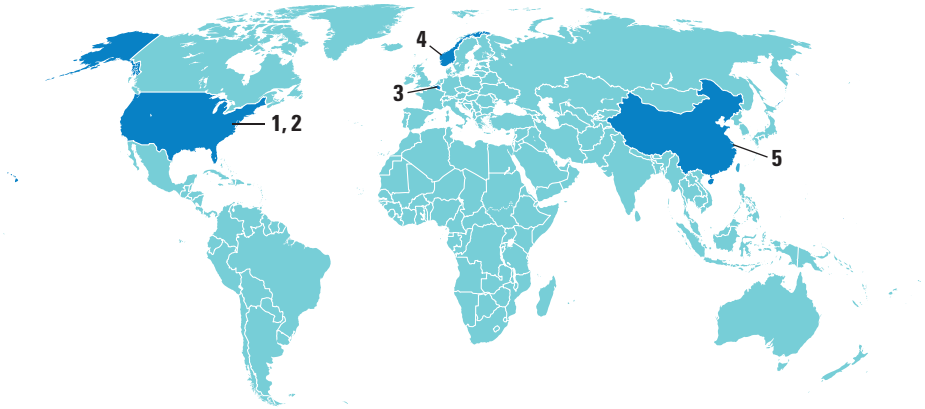
Early prevention of mass violence emphasizes giving help to countries in trouble before violence occurs, if possible. Danger signals are typically evident years before there are casualties, providing ample warning time. But specific response options with contingency plans must be in place to respond to such warnings. To prepare, a critical mass of knowledge and skill is required, embodied by a core of professional scientists, scholars, diplomats, lawyers, and political and military leaders, as well as specialists in the fields of conflict resolution and violence prevention. Within a nation, government-university relations are critical for research, training, and practice. But strong support should come from a network of cooperating regional organizations, in addition to international centers.

The immense danger of mass violence is a powerful stimulus for using our extraordinary learning, problem-solving, and survival capacities to diminish great risks, as was successfully done in the aftermath of the Cold War. Regular meetings over several years between distinguished U.S. and Soviet scientists on arms control, crisis prevention, and conflict resolution helped to reduce the nuclear danger and reconcile grievances. The problems of prejudice, ethnocentrism, hatred, and violence are still too low on the world's priority list. Science can continue to clarify the factors that affect these issues. And it is incumbent on the scientific community to use its deep and broad knowledge to help move conflict prevention to a much higher place on the global agenda.

– David Hamburg

10.1126/science.1223918

AROUND THE WORLD



Washington, D.C. **1**

Senate Bill Would Preserve Helium Supply for Research

A new bill would alter the ongoing sales of the United States' reserve of helium, an element used in the manufacture of semiconductors and microchips and essential for cooling experiments to near absolute zero. The proposed Helium Stewardship Act of 2012 would ensure a supply of helium for science through about 2030 by slowing sales of the reserve, held near Amarillo, Texas (pictured); requiring the federal government to charge the market price for helium; and holding back the last 85 billion liters of the gas for federal users, including holders of research grants. The bill would also avert an acute shortage by continuing the sales past next year, when the 1996 law authorizing them will expire. The sales began in 2003 and supply 30% of the world's helium.

Moses Chan, a physicist at Pennsylvania State University, University Park, welcomes the bill, which was proposed by Senator Jeff Bingaman (D–NM). "They are saying that they are not getting out of the helium business" after next year, says Chan, who served on a 2010 study by the National Academies' National Research Council that criticized the current fixed-price sales. **http://scim.ag/heliumbill**

Washington, D.C. **2**

Summit Generates Global Principles for Peer Review

A group of research administrators from 44 countries has issued the first-ever international principles governing peer review of grant proposals. Meeting this week at the National Science Foundation, the heads of research agencies from around the world pledged to use experts with the appropriate knowledge, disclose in advance any potential conflicts of interest, maintain the confidentiality of a proposal's content, provide applicants with relevant feedback, and make their rules transparent to the community.

Participants also announced the formation of a Global Research Council, a virtual entity that will tackle important problems affecting funding agencies around the world. Regional meetings will provide input for an agreement to be unveiled at an annual gathering held at a different site each year. Next year's meeting will be sponsored by Germany and Brazil, for example, and will take place in May in Berlin. **http://scim.ag/peerrevsummit**

Brussels **3**

European Parliament Rebukes Regulatory Agencies

The European Parliament on 10 May postponed signing off on the 2010 expenditures of three science-related agencies amid concerns over conflicts of interest and inappropriate spending.

The vote to postpone the so-called discharge—which affects the European Food Safety Authority (EFSA), the European Medicines Agency, and the European Environment Agency—has no immediate consequences, but it is seen as a warning that the agencies need to restore their credibility. It came 2 days after EFSA's head, Diána Bánáti, resigned because she had returned to her job at an industry-funded group.

The agencies, which provide the European Commission with scientific advice, have been under fire from members of the European Parliament for several years for being too close to industry and other parties. Members have used their right to withhold approval of past spending as a way to express their discontent and put pressure on the agencies.

The European Parliament will vote on the three budgets again in October; it has asked the agencies to come up with stricter and more transparent conflict-of-interest policies first. **http://scim.ag/sciencerebuke**

Mongstad, Norway **4**

Norway Unveils Carbon Capture Plant

Norway opened the doors to the world's largest carbon capture and storage (CCS) test facility on 7 May. The project, called the Technology Centre Mongstad (TCM), is a joint venture between the Norwegian government, Norwegian energy company Statoil, South Africa's Sasol, and Shell. At $1 billion, TCM's price tag is 10 times the original cost estimate.

TCM's aim is to test, verify, and demonstrate carbon-capture technologies, which currently are expensive and consume a lot of power. The center will test two different liquid-based post-combustion technologies, each designed to filter out as much as 85% of the carbon dioxide in flue gas. One uses chilled ammonia and the other uses an amine solvent to trap the gas. The plant's operators will experiment with different concentrations of carbon dioxide and different flow rates, using the center's more than 4000 instruments to monitor the tests.

The TCM test center can't yet store the carbon that it captures, however. The full-

CREDITS (TOP TO BOTTOM): HELGE HANSEN FOR TCM DA; JOYCE MARSHALL/*FORT WORTH STAR-TELEGRAM*/MCT/LANDOV

Published by AAAS

Gene Flow Makes for Copycat Color Patterns

This colorful butterfly is one of hundreds of subspecies of the 43-species genus *Heliconius*, which has varied color patterns and is considered a classic example of mimicry used to ward off potential predators.

An analysis of the DNA sequence from three *Heliconius* species and several subspecies reveals that those with the same color patterns have the same versions of key genes—holdovers from extensive hybridization within the genus, the authors report this week in *Nature*. Hybridization is thought to be detrimental, as offspring are often less fit, but the acquisition of the right color patterning genes proved advantageous enough that the genes were retained, says Adriana Briscoe, an evolutionary biologist at the University of California, Irvine.

The *Heliconius* genome sequencing project also revealed much greater than expected conservation between moth and butterfly genomes. The butterfly also has an impressive array of olfactory receptor genes. Until now, butterflies were not thought to rely much on a sense of smell, unlike moths, which depend heavily on long-distance odors. But that difference may not hold up; instead, moths may really be "butterflies of the night," says Briscoe.

CREDITS (TOP TO BOTTOM): CHRIS JIGGINS/UNIVERSITY OF CAMBRIDGE; INSTITUT LAUE-LANGEVIN; LBL

scale carbon capture center, including storage, was originally scheduled to be online in 2014. It has been delayed to 2018.

Shanghai, China **5**

U.S. Tweets Chinese Air Quality

A new Twitter feed run by the Shanghai U.S. consulate is charting the city's air quality—and so far the prognosis is not good. On 12 May, the feed, @CGShanghaiAir, began reporting hourly readings of concentrations of harmful particulate matter less than 2.5 micrometers in diameter ($PM_{2.5}$), as measured inside the consulate's compound. The bulk of tweets in the feed's first few days reported air quality index levels from 100 to 200—which the U.S. Environmental Protection Agency classifies as "unhealthy for sensitive groups" or "unhealthy."

The consulate's reports conflict with those by the Chinese government, which have largely focused on measurements of PM_{10} (*Science*, 1 August 2008, p. 636) and the new feed could rile Chinese officials, who protested similar air quality tweets by the U.S. Embassy in Beijing in 2009. After locals disseminated the Beijing embassy results on Chinese microblogs, China took some steps in January toward openness. But more accurate figures won't fix the air itself. And that, as @BeijingAir reported on a day when index levels exceeded 500, can be "crazy bad."

NEWSMAKERS

Senior DOE Official Resigns

Arunava Majumdar, the head of the Department of Energy's Advanced Research Projects Agency-Energy (ARPA-E), will leave his post 9 June, Energy Secretary Steven Chu wrote in a 9 May e-mail to agency staff members. Biochemist Eric Toone, a former professor at Duke University in Durham, North Carolina, and the deputy director of technology for ARPA-E, will become the new head.

Majumdar

Majumdar's departure "is a kick in the stomach," but Toone "will keep the agency in good hands," says Barton Gordon, a former member of the U.S. House of Representatives who spearheaded the creation of ARPA-E in 2007. Gordon, now a lobbyist with K&L Gates in Washington, D.C., says Majumdar is a "good scientist and a good organizer who created a good bipartisan following [in Congress] for ARPA-E."

Modeled after the Pentagon's Defense Advanced Research Projects Agency, ARPA-E is designed to funnel money quickly to high-risk, "transformational" efforts to develop new energy technologies. The agency has won broad backing from industry, and has fared relatively well in the annual budget battles. This year it will spend about $300 million on a wide array of projects, down from a high of $400 million in 2010. **http://scim.ag/Majumdar**

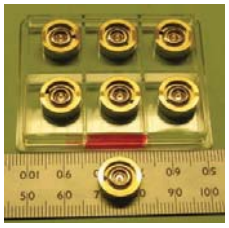
THEY SAID IT

"Industry has used this study in ways that are improper and untruthful."

—Fire protection engineer Vytenis Babrauskas of Fire Science and Technology Inc. in Issaquah, Washington, commenting to the *Chicago Tribune* on 9 May about how chemicals manufacturers have allegedly misrepresented a study he conducted 25 years ago while he was at the then–National Bureau of Standards. The study examined whether flame retardants added to furniture and other household objects could slow fires.

FINDINGS

Gamma Ray Bending Opens New Door for Optics



Physicists have created a lens for highly energetic light known as gamma rays. Once thought impossible, the feat may open up a new field of gamma-ray optics for medical imaging, detecting illicit nuclear material, and getting rid of nuclear waste.

The electrons in glass lenses >>

Science LIVE

Join us Thursday, 24 May, at 3 p.m. EDT for a live chat on nanomedicine. http://scim.ag/science-live

Published by AAAS

BY THE NUMBERS

64% Percentage of global deaths of children under 5 caused by preventable infectious diseases in 2010, according to a study published on 9 May by *The Lancet*.

3.6% The percentage of adults in the United States who sleepwalk, according to a study published on 14 May in *Neurology*. The study looked at 19,000 adults in 15 states.

1,000,000 Number of pages now contained in the online biodiversity resource Encyclopedia of Life, which reached the milestone on 9 May.

Random Sample

Mastering Clairvoyance

A recent master's degree thesis at European University Viadrina Frankfurt (Oder) in Germany is generating waves—though not the "time waves" the student was looking for.

Orthopedist Peter Conrad had set out to study a hollow aluminum cylinder known as a Kozyrev mirror, which proponents claim opens a spacetime channel that enables a person to communicate telepathically, see aliens and UFOs, and even contact the dead.

In what thesis supervisor Harald Walach of the Institute for Transcultural Health Studies at Oder calls a "smart experiment," Conrad tested whether the mirror also facilitated clairvoyance. Test subjects held on to wires either connected to a small Kozyrev mirror or to a mockup. A piece of paper with a number written on it was placed in the cylinders. In one set of experiments, he found, people connected to a Kozyrev mirror were significantly better than average at guessing the numbers.

The work won Conrad his master's degree last year—but some scientists see it as part of a worrying trend of pseudoscience finding a home in publicly funded universities. Responding to those critics, Walach wrote in a statement: "This is an exploratory finding, that can be seen as a sign for clairvoyance."

Walach's institute has sparked debate before when it included energy medicine, faith healing, and other esoteric practices on its curriculum in 2010. Those plans were later scrapped. But there have been similar situations at other universities in recent years. "It appears that such nonsense clusters at certain institutes and around certain professors," says German-born Edzard Ernst, professor of complementary medicine at the University of Exeter in the United Kingdom. "It is regrettable and alarming when German universities produce such hogwash disguised as science."

>>FINDINGS

that refract visible light are less able to bend incoming light at ultraviolet frequencies. Physicists have made lenses for x-rays by stacking together layers of patterned material, ultimately allowing imaging at a nanoscale resolution.

Now, a team of physicists led by Dietrich Habs at the Ludwig Maximilian University of Munich in Germany and Michael Jentschel at the Institut Laue-Langevin in Grenoble, France, has found a way to bend gamma rays, too. They directed gamma rays into a crystal spectrometer, which sent half through a silicon prism and into another spectrometer to measure their final direction; the other half went straight to the spectrometer. Gamma rays with an energy above 700 kiloelectronvolts were slightly bent by the silicon prism, the team reports in a paper to be published this month in *Physical Review Letters*. **http://scim.ag/gammarayoptics**

New Candidate for World's Oldest Cave Art

Since their discovery in 1994, the 37,000-year-old paintings of lions, rhinos, and other animals in southern France's Chauvet Cave have been the oldest known cave art. Now a team working at the Abri Castanet, a cave in southern France's Vezere valley, claims to have discovered evidence of artistic activity that may be older than Chauvet.

Working in the cave, archaeologist Randall White of New York University in New York City and his team found limestone blocks adorned with engravings, including one the team interprets as female genitalia. The blocks were impossible to date, but in 2007, the team excavated another fallen block engraved with a vulva-like image; the block rested directly on a segment of cave once occupied by humans and littered with animal bones that dated to between 36,000 and 37,000 years ago.

That would mean that the works of art at Abri Castanet are at least as old as those at Chauvet, White and colleagues conclude in a paper published online 14 May in the *Proceedings of the National Academy of Sciences*. http://scim.ag/Abricaveart

CREDIT: GIGAPAN.COM/GIGAPANS/56492

Published by AAAS

At odds. Al Gilman of CPRIT and Lynda Chin of MD Anderson (*left*) disagreed on a review of her $18 million drug project.

BIOMEDICINE

A Texas Wrangle Over Cancer Research Funds

On the way to becoming a world-class funding resource, Texas's 3-year-old, $3 billion cancer research agency has run into management troubles. Last week, Alfred Gilman, the Nobel Prize–winning biochemist who serves as the chief scientific officer of the Cancer Prevention and Research Institute of Texas announced he is stepping down this fall because he believes CPRIT's leaders are making research funding decisions that bypass scientific review. Gilman's concerns have rattled CPRIT's board members. They are now scrambling to reassure reviewers and the public—including Texas lawmakers who voted to fund CPRIT—that they remain committed to peer-reviewed science.

Gilman says his resignation was triggered in part by a controversial "incubator" grant of $20 million awarded in March. Most of the money will go to a drug-discovery institute that Lynda Chin helped launch when her husband Ronald DePinho became president of the University of Texas (UT) MD Anderson Cancer Center in Houston. Gilman claims the Chin grant application "appears to have been glued onto" another proposal from Rice University and reviewed by business experts, not scientific reviewers, "unusually quickly." CPRIT officials and Chin deny this, saying it boils down to a difference of opinion about the appropriate form of review.

Gilman has another complaint: He claims that pressure from some members of CPRIT's oversight committee—appointed by the governor and other elected officials—caused CPRIT's director to put a hold on grants worth $39 million in March because most of the money was going to Gilman's former institution, UT Southwestern Medical Center in Dallas. Some board members think more CPRIT funding should be going to their local institutions, Gilman says. The eight members of CPRIT's scientific review council, an advisory group headed by fellow Nobelist and Massachusetts Institute of Technology biologist Phillip Sharp, wrote the board this week to say that they share Gilman's concerns. "Here is an explicit example of how peer review runs up against specific interests," Sharp says.

Others suggest that the dispute arose in part from tensions within CPRIT about the balance between commercialization programs, which make up 17% of its grants portfolio, versus basic research, which makes up 73%. (By law, another 10% must be spent on cancer prevention.)

CPRIT sprang to life 5 years ago when Texas voters approved a plan to fund the agency with bond sales amounting to $3 billion over 10 years, following a model similar to California's stem cell research agency. Initially, some worried about political influence, but the concern faded when Gilman joined the agency. The peer-review system he created includes more than 100 scientists from outside Texas. CPRIT has disbursed $500 million so far for 387 research grants and recruited nearly 40 researchers to the state (*Science*, 27 May 2011, p. 1019).

Chin and DePinho resigned from the Dana-Farber Cancer Institute in Boston and moved to Texas in August. They brought about 20 staff scientists from the Belfer Institute for Applied Cancer Science, an academic-industry drug discovery program they founded at Dana-Farber in 2004. At MD Anderson, they created the Institute for Applied Cancer Science (IACS) with Chin as scientific director. It's "unique," Chin says, because scientists discover and develop new drugs themselves and can work on a dozen or more projects. They make "fast kill" decisions when appropriate. MD Anderson pledged to fund IACS at $15 million a year for 5 years.

Chin says soon after she arrived in Texas, CPRIT officials encouraged her to apply for an "incubator" grant designed for "the discovery or development of innovative oncology medicines." CPRIT "approached us," Chin says. "They believed this is exactly what an incubator should do." In accordance with the request for applications (RFA) approved by CPRIT's board, she says, her team began preparing a "business plan" that would not go through scientific review.

Separately, last November, CPRIT's commercialization reviewers approved an incubator proposed by Rice University to set up a center to offer expert advice to Houston-area researchers seeking to commercialize discoveries. But the reviewers held the proposal until Rice had a management team, says CPRIT commercialization chief officer Jerry Cobbs. He decided that combining it with Chin's IACS center "would be a much more powerful tool completing the innovation pipeline."

Cobbs says he had Chin present IACS's plans to Gilman by videoconference in January to get his "insight." Gilman expressed concerns: To him, IACS appeared to be a research program because its work included basic biomedical and chemistry studies, and it did not identify products or a company.

CREDITS (TOP TO BOTTOM): ZERESHK/WIKIMEDIA COMMONS; UT SOUTHWESTERN MEDICAL CENTER; F. CARTER SMITH/MD ANDERSON CANCER CENTER

Published by AAAS

He urged Chin to submit her proposal as a research grant. Chin disagreed: "This is an industry-like entity," she says. A scientific review would "not be appropriate" because IACS staff members are former industry scientists and can't be evaluated as academics are, such as by reviewing their publications, she claims.

CPRIT Executive Director William Gimson says he decided Chin's proposal should be submitted as an incubator despite Gilman's objections. Chin submitted a seven-page proposal on 11 March; the commercialization panel considered it 10 days later, issuing four written reviews, Chin says. The executive board approved the $20 million award—its largest single-year award ever—at its 29 March meeting. Between $15 million and $18 million will go to MD Anderson, Gimson says.

The quick review simply shows that "we're nimble, we're entrepreneurial," Gimson says. If the proposal had not been "compatible with the RFA," Gimson says, his commercialization reviewers—venture capitalists and biotechnology experts, many with scientific expertise—would have rejected it. To Gilman and other academic scientists, however, the MD Anderson award amounts to a clever run around the scientific review process. In his 8 May resignation letter, Gilman objects that CPRIT has awarded "vast funds for research programs" that were not adequately described and therefore "could not have been reviewed."

Adding to his concerns, Gilman says, Gimson did not bring to the board's March meeting seven multi-investigator grants approved by the scientific review council. Sharp and other council members write in their letter that Gimson worried that some board members would block the well-reviewed grants because most of the funding was going to UT Southwestern. This suggests "bias," Sharp and the councilors wrote, which "we vigorously deny."

Gimson has a different explanation: He says he held the grants because the board has been keen to fund more commercialization projects and, in any case, the total funds he had available weren't sufficient to cover the $20 million incubator award, other peer-reviewed grants, and the $39 million required for the seven projects approved by the scientific council. He expects to bring the grants to the board in July, he says.

The former director of the Fox Chase Cancer Center in Philadelphia, Pennsylvania, Robert Young, a member of CPRIT's commercialization review council, says he recused himself from the IACS review because he serves on the board of a company DePinho and Chin founded. Still, he says the center they're planning in Texas is "a proven model" that might look "unconventional from the point of view of a scientific review committee." He says it's not the first time the commercialization and scientific branches of CPRIT have differed on proposals.

Gimson says CPRIT is taking steps to ensure that "this doesn't happen again." The board may clarify the "line" between an incubator and academic research project and in July may consider whether to set aside a fixed amount of its budget for commercialization. Gilman says he would be supportive, but that balance "isn't the issue." It is "disguising" a large research grant as commercialization to "bypass peer review."

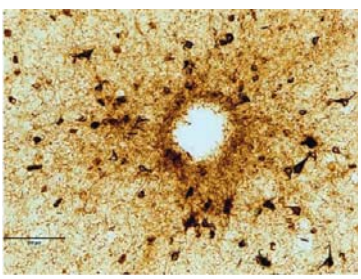
Gilman, 70, plans to retire on 12 October after a 42-year scientific career. He hopes to stay at CPRIT until then in part to make sure applicants "will still encounter a functional peer-review system," he says in his letter. Sharp says he and other reviewers will be closely watching the board's actions in the coming months. **–JOCELYN KAISER**

NEUROPATHOLOGY

Blast Injuries Linked to Neurodegeneration in Veterans

As the number of troops exposed to improvised explosive devices (IEDs) in Iraq and Afghanistan has mounted, so, too, has concern about the long-term impact of these and other blasts on the neurological health of service members. These worries are amplified by recent evidence that head injuries sustained by football players and other athletes can cause personality changes, dementia, and neurodegeneration later in life. A new study ties these troublesome threads together: In autopsies of four military veterans who served in recent conflicts, researchers have found distinctive features of the same neurodegenerative disease reported previously in athletes.

The team, led by Lee Goldstein, a physician-scientist who focuses on neurodegenerative disease at Boston University, and Ann McKee, a neuropathologist at the Bedford Veterans Affairs Medical Center in Massachusetts, reports its findings this week in *Science Translational Medicine* (http://scim.ag/LEGoldstein). Goldstein, McKee, and 33 collaborators with expertise in areas including blast physics and neurobiology also describe a mouse model of blast injury that captures several features of the neuropathology seen in the military veterans and athletes and provides preliminary clues about the underlying mechanisms.



Dark omen. Pathological tangles of tau protein (dark areas) have been found in brain tissue from four veterans.

CREDIT: L. E. GOLDSTEIN *ET AL.*, *SCIENCE TRANSLATIONAL MEDICINE* **4**, 134 (16 MAY 2012)

In the past 10 years, the widely reported suicides and accidental deaths of athletes have sparked inquiries into whether even seemingly minor blows to the head can have lifelong consequences, especially when they occur repeatedly (*Science*, 7 August 2009, p. 670). Autopsies of dozens of former players have revealed a condition that has come to be known as chronic traumatic encephalopathy (CTE). Its hallmark is the abnormal accumulation of a protein called tau, which also forms pathological tangles in people with Alzheimer's disease and frontotemporal dementia. Many of the athletes diagnosed with CTE on autopsy had a history of problems with anger, rash and risky decision-making, impairments of memory and attention, and alcohol or drug abuse.

Clinicians and researchers working with troops returning from Iraq and Afghanistan have been seeing some similar symptoms (*Science*, 29 July 2011, p. 514). But until recently, there were no published cases of CTE in veterans confirmed by autopsy, currently the only definitive test. That changed with a November 2011 report in *Neurosurgical Focus* by Bennet Omalu and colleagues. Omalu, the chief medical examiner in San Joaquin County, California, was also the first to report CTE in a professional football player in 2005. His November report describes the case of a Marine who died by suicide at age 27 after two tours in Iraq during which he had two close encounters with mortar and IED blasts. After his return, the Marine was diagnosed with post-traumatic stress disorder; he

Published by AAAS

Dangerous duty. The long-term health impacts of IED blasts aren't yet known, but a new study suggests a worrisome link to neurodegenerative disease.

CREDIT: DAVID GOLDMAN/AP

became forgetful and irritable and drank more heavily. Omalu and colleagues found a pattern of tau tangles in his brain that is characteristic of CTE.

McKee's group has now contributed four new cases to this grim literature. These veterans, men between the ages of 22 and 45, suffered various combinations of cognitive, emotional, and impulse-control problems before dying of suicide and other causes. Postmortem exams revealed tangles of tau and damage to axons, the spindly extensions that convey signals from one neuron to another. In athletes with CTE, these abnormalities occur primarily in the frontal lobes of the brain, McKee says. In the blast-injured veterans, the pathology is more evenly distributed. "But when you look microscopically, it's very, very similar," she says.

"Whether there is a relationship between CTE and blast … is definitely a question that we feel needs to be answered," says Colonel Dallas Hack, a physician who coordinates trauma research for the U.S. Armed Forces. In his view, the new cases are suggestive but not conclusive. He notes that three of the four cases had previous concussions from sports, fights, or accidents, and one was not exposed to a blast in combat. (Omalu's case also had a history of concussion in civilian life.) "Was their military service part of the cause? We don't know that yet," Hack says.

The mouse experiments presented in the new paper bolster the case for causation because the mice were exposed to a blast and nothing else, says co-author Goldstein. The researchers placed mice in a 4.5-meter-long tube and used compressed gas to create a shock wave comparable to what a soldier might experience from an IED blast at close range. The setup differed from previous blast-tube experiments in that the animals' heads were unrestrained. This more closely mimics what happens in combat by allowing the so-called bobblehead effect, in which a blast momentarily causes the head to flop back and forth on the neck, Goldstein says.

In these experiments, 2 weeks after exposure to a single blast, mice exhibited accumulations of chemically modified, or phosphorylated, tau in their brains, suggesting that the pathological process that leads to the fully formed tangles seen in the veterans' brains was already under way. (Tau occurs naturally in neurons, but phosphorylated forms of the protein accumulate to form pathological tangles.) Blast-exposed mice also exhibited inflammation and damage to axons and capillaries in the brain.

The researchers also found sick and dying neurons in the hippocampus, a brain region important for learning and memory. Given the common complaints of forgetfulness and other cognitive problems in veterans, Goldstein and colleagues followed up with additional experiments that revealed abnormalities in the electrical signaling of hippocampal neurons in blast-exposed mice, as well as learning and memory deficits on a maze test. These deficits did not occur when the researchers immobilized the head of a mouse during a blast, however, which suggests to Goldstein that it's the whipsaw motion of the head that initiates the injury.

"This is a really good piece of work," says David Hovda, a neuroscientist who studies brain injury at the University of California, Los Angeles. Researchers have long debated whether the biomechanics of blast injuries are distinct from those of other types of head injuries, Hovda says. The new study, which includes high-speed video and pressure measurements during blast exposure, argues they are not, supporting the theory that blasts injure the brain by causing acceleration and deceleration of the head, Hovda says: "That would mean we're looking at the same type of injury in Afghanistan that we're looking at on the football field or in a car accident."

David Brody, a neurologist at Washington University School of Medicine in St. Louis, is intrigued by the axon damage in the veterans' brains. Last year, his group reported in *The New England Journal of Medicine* evidence of axon damage in military personnel using a type of magnetic resonance imaging called diffusion tensor imaging. Brody would like to see the Boston researchers create more detailed maps of axon damage that could be compared with his group's imaging data. The mouse model, he adds, should be useful for investigating the mechanisms of axon injury, as well as the important question of whether tau is a cause or effect of neural injury.

Either way, the findings suggest that phosphorylated tau could be a useful biomarker for blast injury, says David Menon, a neurointensive care physician at the University of Cambridge in the United Kingdom. "The question, then, is how you … detect the biomarker in vivo rather than having to wait for patients to die," Menon says. In the past decade, researchers have developed diagnostic brain scans to detect amyloid, a peptide associated with Alzheimer's disease, but similar scans for tau have lagged behind.

Going forward, a major question for both athletes and military personnel is the number and type of hits it takes to cause CTE, says Daniel Perl, a neuropathologist at the Uniformed Services University of the Health Sciences in Bethesda, Maryland. Perl has funding from the Army to create a brain bank repository he hopes will enable researchers to connect the dots between head injuries (gleaned from military service records), clinical problems (from medical records), and neuropathology (from autopsied brains). Omalu and neurosurgeon Julian Bailes of NorthShore University HealthSystem in Chicago have co-founded an organization called the Brain Injury Research Institute that also aims to examine the brains of veterans for evidence of CTE via collaborations with medical examiners around the country. It remains to be seen whether CTE will become a bugbear for veterans as it has for athletes.

–GREG MILLER

Published by AAAS

AIDS RESEARCH

FDA Panel Recommends Anti-HIV Drug for Prevention

SILVER SPRING, MARYLAND—On 10 May, the Antiviral Drugs Advisory Committee of the U.S. Food and Drug Administration (FDA) held a marathon debate about whether an anti-HIV drug on the market as a treatment should receive approval as a preventive for uninfected people. For more than 12 hours, the committee heard scientific evidence and impassioned arguments for and against, ultimately recommending that FDA approve the use of the drug Truvada for what's called pre-exposure prophylaxis (PrEP). The decision was not unanimous, and there was a protracted back and forth about how to reduce the possibility that PrEP might cause more harm than good. By the time the committee chair asked whether the 22 members were ready to vote—which took place after the scheduled 6:30 p.m. adjournment—one person in the audience said, "Amen!"

Minor discord. Michael Weinstein of the AIDS Healthcare Foundation led a small but vocal campaign against the approval of Truvada as a preventive.

There's little question that Truvada, made by Gilead Sciences Inc. in Foster City, California, can prevent sexual transmission of HIV: Large, controlled studies in both uninfected men who have sex with men (MSM) and uninfected heterosexuals who have long-term partners known to be infected have proved that the drug reduces risk by more than 90% when taken daily. But adherence is the rub. Many of the participants in clinical trials had difficulty taking the pill each day, dramatically reducing the overall efficacy in the two pivotal studies with positive results—and leading to outright failure in two large trials in heterosexual women (*Science*, 16 March, p. 1291). This intermittent use raises a host of troubling questions, including whether PrEP will encourage risky behavior or fuel drug resistance to Truvada, outcomes that could seriously undermine its benefits.

There's widespread agreement that current prevention efforts fall short. For the past 2 decades, new HIV infections in the United States have remained steady at roughly 50,000 people a year. Susan Buchbinder, an epidemiologist at the San Francisco Department of Public Health, explained to the committee that the majority of these infections have occurred in MSM; incidence in this group rose from 2006 to 2009, despite intensive efforts to counsel people about reducing their risks.

A central concern about PrEP, shared by advocates and opponents alike, is that it will lead to "risk compensation"—in other words, people will assume the pill protects them and abandon other proven prevention strategies like condom use. A small PrEP study Buchbinder participated in extensively evaluated risk compensation and found that none occurred. Similarly, none of the large-scale PrEP studies found increases in risky behavior, she said, noting that other fields assess risk compensation differently. "We're not asking whether people who are on statins are eating more ice cream," said Buchbinder, who supports Gilead's request for a label change to indicate that Truvada can be prescribed either to treat or prevent infections.

During the public comment period, two dozen staff members from the AIDS Healthcare Foundation—which bills itself as the largest AIDS organization in the world—waged a coordinated attack against the label change, with several insisting that risk compensation will occur. Michael Weinstein, head of the foundation, argued that people who took the pill would assume they didn't need to use condoms. "If you want to do no harm, do not reduce the use of condoms," he said.

The advisory committee spent much of the day wrestling with the possibility that PrEP could create massive drug resistance to Truvada, a mainstay of antiretroviral treatment. "The potential harm here is stupendous," said the committee's acting chair, Judith Feinberg of the University of Cincinnati College of Medicine in Ohio.

Although Truvada, a combination of the antiretrovirals tenofovir and FTC, works well by itself to prevent an HIV infection, as a treatment, it must be used with other drugs to avoid the emergence of resistant strains. The challenge with PrEP, then, is making sure that people use Truvada as a solo drug only if they are not infected—otherwise, resistant strains could run rampant and render the drug useless as both a treatment and a preventive.

If FDA approves the label change, Gilead explained in a "risk mitigation" plan how it would educate providers about the importance of prescribing PrEP only to patients who test HIV negative. But making sure only uninfected people use PrEP is easier said than done. In clinical trials of PrEP, few cases of drug resistance were seen in the thousands of study participants. But researchers checked for infection each month and also ran sensitive PCR assays that can detect HIV in the first few weeks after infection, which is missed on standard antibody tests. In the real world, who would oversee repeated tests of people prescribed PrEP? Would retesting be required for refills? It is possible to restrict access to drugs—women receiving the acne medication Accutane must receive pregnancy tests before each prescription is filled—but as several committee members stressed, it would create a huge burden to do something similar with PrEP.

The final vote recommended that FDA approve Gilead's request without a testing requirement. The decision outraged Weinstein. "It's a very, very sad day in the battle against AIDS in America," he complained. If FDA accepts the committee's recommendation—which is expected—Weinstein said it would be a "reckless act" and a "new Tuskegee experiment."

But Weinstein and his group clearly were the minority voice. "The meeting and vote represent a tremendous milestone for HIV prevention," said Jared Baeten, an epidemiologist at the University of Washington, Seattle. And Baeten, along with the committee, said many of the questions surrounding PrEP can best be answered by carefully studying what happens after approval.

–JON COHEN

CREDITS (LEFT TO RIGHT): J. COHEN/*SCIENCE*; COURTESY GILEAD SCIENCES INC.

Published by AAAS

NEWSMAKER INTERVEW: STEPHEN PRUITT

Coalition Begins Push for U.S. Schools To Adopt Voluntary Science Standards

Herding cats is a cinch compared to what Stephen Pruitt is trying to do: coax as many states as possible into adopting a set of standards for teaching science in their elementary and secondary schools. The first public draft of what are being called the Next Generation Science Standards appeared last week, and Pruitt has made his job even harder by inviting comments on their contents (nextgenscience.org).

Pruitt works for Achieve Inc., a coalition of high-tech companies, foundations, and state and local governments that hope to use their collective influence to create a voluntary national science curriculum where none now exists. The federal government is not a partner because what happens inside U.S. classrooms is left to the discretion of each state and some 14,000 local school districts. In recognition of that decentralized structure, Achieve has signed up 26 "lead" states to help write the standards and to be cheerleaders for their adoption.

How hard could that be? The new standards try to streamline the curriculum by highlighting the most important concepts and encouraging teachers to find ways for students to acquire a richer understanding of them. The biggest challenge for educators will be becoming comfortable with the standards' novel approach of linking core ideas in the physical, life, earth, and engineering sciences to what are called "science and engineering practices," that is, anchoring them in how scientists tackle a problem. A document issued last summer creates a "framework" on which to hang the new standards (*Science*, 29 July 2011, p. 510). Each lesson is also tied to its real-world application, as well as to the content of a separate effort embraced by 46 states that established "common core" standards in reading and mathematics.

But the science standards don't exist in a political vacuum. When Pruitt says, "We haven't created any new science," he's referring obliquely to the controversy over teaching evolution and climate change in U.S. schools. Pruitt is crossing his fingers that those never-ending debates won't derail the new science standards.

The Carnegie Corporation of New York has provided nearly $3 million for the project, which Pruitt hopes will be completed by March 2013. That timetable allows for circulation this fall of a second draft reflecting any changes from the initial round of comments.

A former high school chemistry teacher and state administrator in Georgia, Pruitt offers a tutorial on the new standards in this edited interview with *Science*.

–JEFFREY MERVIS

CREDIT: COURTESY OF ACHIEVE INC.

> *"With the success of the common core in reading and math, states are starting to see the benefit of having a common set of standards. … Standards are as much about equity as anything."*
>
> —STEPHEN PRUITT, ACHIEVE INC.

Q: Why should researchers care about the new draft standards?
S.P.: They represent what students will need to know to be college ready. They also give the community an opportunity to help refill the STEM [science, technology, engineering, and mathematics] pipeline.

Q: Why haven't we had standards before?
S.P.: There have been standards, including the Benchmarks for Science Literacy by AAAS [*Science*'s publisher] and the National Science Education Standards from the National Academies. But they were never meant to be adopted by states as standards. Real standards must include student expectations, and states have different levels of expectations.

Q: Why have standards at all?
S.P.: With the success of the common core in reading and math, states are starting to see the benefit of having a common set of standards both for kids moving across state lines and to ensure that all kids are getting equal access to a good science program. In my opinion, standards are as much about equity as anything.

Q: The Achieve Web site says that "lead states" need only "give serious consideration" to adopting the standards once they have been finalized. Why so permissive?
S.P.: We can't expect a state to adopt these sight unseen. Also, states with strong science standards will want to make sure these are at least as good as what they already have. At the same time, we required the lead states to convene broad-based groups of stakeholders to help write the standards in hopes that would facilitate their adoption.

Q: How much do these standards look like what already exists, for example, in Georgia?
S.P.: We didn't create any science. The thing that is different is this integration of disciplinary content and crosscutting practices. When you look at the verbs, they are not the traditional "explain and analyze." These are more like "construct a model to explain" or "develop an argument that justifies the claim that." And that's a pretty significant change.

So the content may be similar, but they will be different than existing standards in most states.

Q: Do you expect states to try to revise them?
S.P.: We actually don't want states to unpack the standards. The foundation boxes are there to provide specifics of what it means for middle school students to construct a model, for example, in a lesson on structure and patterns of matter. It's new, so there will need to be some training. Teachers might get a little nervous at first because it looks more complex. But then they realize that it actually gives them a lot more information about what to do.

Q: Will the Next Generation Science Standards be a failure if states do unpack them?
S.P.: I think it's too early to talk about failure. We need to get through the second draft before we address that question.

Published by AAAS

PUBLIC HEALTH

China Takes Aim at Rampant Antibiotic Resistance

SHANGHAI—In a true-life drama on the Chinese talk show *Shihou Zhu Geliang*, a young man has so little confidence in the hygiene of his company's cafeteria that he supplements each meal with antibiotics. That is his undoing. When the man later falls ill, drugs fail to save him. An autopsy reveals why: His body is riddled with multiple strains of drug-resistant bacteria.

The case may have been extreme, but *Shihou Zhu Geliang* had a serious message. The episode accompanied a government-led crusade to warn Chinese against the perils of frivolous antibiotic consumption. The campaign culminated last week in a Health Ministry directive laying out stricter regulations for prescription drugs. "Problems with public awareness and with doctors' [prescribing] habits have led to excessive antibiotic use in China," says Ni Yuxing, head of the clinical microbiology department at Ruijin Hospital here. "And that has created resistant bacteria."

Bacteria that cannot be stopped by common drugs are proliferating around the world (*Science*, 18 July 2008, p. 356). But a health care system that encourages doctors to churn out prescriptions, intensive marketing by pharmaceutical companies, and heavy use of antibiotics in animal husbandry and fisheries make China a special case. More than 60% of *Staphylococcus aureus* isolates from Chinese patients in surveyed hospitals in 2009 were methicillin-resistant—the dreaded MRSA—up from 40% in 2000. The proportion of *Streptococcus pneumoniae* isolates resistant to macrolides, meanwhile, now tops 70%. Roughly the same share of *Escherichia coli* isolates are resistant to quinolones—the highest rate in the world.

China's health ministry hopes to ward off calamitous outbreaks of drug-resistant strains. The new regulations, which take effect 1 August, are the latest in a string of measures that started with the launch of a drug-resistance-monitoring network in 2004 and an ambitious 3-year program, now in progress, to combat the overprescription of drugs. But critics say that without an overhaul of the health care system, the new measures may have limited success.

"Antibiotic resistance is a serious public health threat in China," says Xiao Yonghong, an infectious disease specialist at Zhejiang University in Hangzhou. He oversees the Health Ministry's National Antibacterial Resistance Investigation Net, which covers 80 hospitals nationwide. Drug resistance is most acute in densely populated cities in the east. Erythromycin-resistant *S. pneumoniae*, for example, appeared in 94% of isolates from children tested in hospitals here in 2004 and 2005. The sole strains of common drug-resistant bacteria not thriving in China are vancomycin-resistant *Enterococci*, Xiao says: "That's the only good news."

China's woes are in part the consequence of earlier health care reforms. Until the early 1980s, China had government-provisioned care bolstered by "barefoot doctors": minimally trained health workers who ministered to patients in remote areas. But near-universal care was ultimately scrapped in favor of a free-market approach. Hospitals needed new revenue streams.

On the drip. In China, antibiotics are commonly taken intravenously for colds and other maladies that are often treated less aggressively in the west.

A burgeoning drug industry came to the rescue. The government allowed hospitals to skim a 15% profit from pharmaceutical sales, and doctors' pay was soon linked to sales. Overprescription became rampant. "The whole thing really boils down to perverse incentives," says Lucy Reynolds, a public health expert at the London School of Hygiene and Tropical Medicine's European Centre on Health of Societies in Transition. By the 1990s, Xiao, who then worked in a hospital in Chongqing, noticed that something was awry: "Patients weren't responding to treatment."

In the meantime, growing numbers of Chinese farmers discovered that rearing livestock on antibiotics yielded larger animals and boosted profits. In a 2007 survey, Xiao estimated that nearly half of the 210,000 tons of antibiotics produced in China end up in animal feed. The U.S. Food and Drug Administration warned in 2007 that Chinese farm-raised fish are laced with fluoroquinolones and other antibiotics not approved for use in U.S. seafood.

Those drugs find their way into the human gut. In an unpublished study, Zhu Baoli of the Institute of Microbiology in Beijing and colleagues sequenced gut microbes in Chinese, Danish, and Spanish people. Chinese guts had the highest number of antibiotic-resistant genes. They were also dominated by genes resistant to tetracycline, which in China is mostly used in animal feed. While that might not affect clinical treatment with drugs other than tetracycline, Zhu says, "some drugs that are used in animals are also used in humans."

China's woes extend beyond the doctor's office and the dinner table. As the story of the man worried about cafeteria food shows, some Chinese pop antibiotics the way Americans pop vitamins. Until 2004, antibiotics were legally available over the counter in China, and families sometimes kept a stash at home.

In 2010, the Health Ministry separated doctors' pay from prescription drug sales. Its new directive goes further, by requiring that drugs be divided into three classes, with drugs with the highest resistance rates to be prescribed only by specialists. Violators can lose prescription rights or their medical license, while offending hospitals can be fined. But the government has not offered hospitals an alternative source of funding to replace drug profits. "A comprehensive overhaul" is needed, Reynolds says.

Even with further reforms, significant obstacles remain. In a survey in Guizhou Province in southwest China, Reynolds found that many doctors mistakenly believed their patients—not the bacteria inhabiting their bodies—had developed antibiotic resistance. Meanwhile, resistance is climbing. Says Zhu: "You now have genes resistant to almost every antibiotic available on the market."

–MARA HVISTENDAHL

CREDIT: LIU JIN/GETTY IMAGES

Published by AAAS

Near Eastern Archaeology Works To Dig Out of a Crisis

In the wake of the Arab Spring, archaeologists in the Near East are locked in a struggle for the survival of their field

New digs. Many regions of the Near East are closed to archaeologists, but new areas, such as this dig at Abu Tbeirah in southern Iraq, are opening up.

WARSAW—The day he left Switzerland for Syria in spring 2011, archaeologist Oskar Kaelin of the University of Basel received a warning from the Swiss foreign ministry about deteriorating political conditions there. He decided to go anyway. It was only his second season leading a team digging a palace left by the mysterious Mitanni Empire, which ruled a large part of the Near East around 1400 B.C.E. "We had the funding, and our luggage was already packed," Kaelin recalls. The team members spent a fruitful 2 months excavating, and when they left in early June, they were among the last foreigners on Syrian soil. They left behind not just their site but also artifacts and samples still awaiting analysis. Kaelin doesn't know when—or if—he can return to complete his research.

The Near East was the birthplace of farming, animal domestication, cities, empires, and writing, and so it exerts a powerful pull on archaeologists. The region is also the birthplace of modern archaeology itself, which began in the 19th century amid the mounds of Mesopotamia. Yet at the field's first gathering* since the Arab Spring unleashed unrest from Morocco to Oman, researchers were worried about the future of archaeology's flagship subdiscipline. More than 120 foreign teams were abruptly shut out of Syria, Egypt is taking a xenophobic turn, parts of Iraq remain prone to violence, and Iran remains virtually sealed off. Even peaceful countries are more difficult for foreign archaeologists to access.

Although about 600 scientists crowded the halls here to listen to nearly 500 presentations, there was a note of quiet desperation in the air. "We're still in shock," says Daniele Morandi Bonacossi of the University of Udine in Italy, who has centered his professional life on the ancient western Syrian capital of Qatna, to which he cannot return for the foreseeable future. "It's getting hard to be a Near Eastern archaeologist," he says. Frank Hole, an emeritus Yale University anthropologist with more than 40 years of experience in the region, agrees: "The discipline is in crisis."

That crisis is due not just to the immediate upheavals. After 3 decades of political instability, grant money is drying up, new students are wary of entering the field, and retiring professors are often not being replaced. But resourceful researchers are finding new places and ways to gather their data. Some have started work in a few stable and long-neglected pockets in the region, such as Iraq's Kurdistan. Others are pioneering advanced remote-sensing techniques, as well as DNA, isotopic, and other analyses to squeeze out more information from the material they can access. As a result, some see a silver lining in the current crisis. "This may prove a moment of promise," says Jason Ur, a Harvard University archaeologist.

Modern archaeology has its roots deep in the dry soil of the Middle East. However, Iran's 1979 revolution, its subsequent war with Iraq, and two gulf wars ended the era of foreign-dominated teams conducting large digs in those two countries. Archaeologists shifted to Syria, Jordan, and Turkey. But Jordan lacks an extensive urban legacy, and Turkey has become increasingly reluctant to grant foreigners dig permits. That makes Syria's closure a hard blow. "Now we are all trying to find other places," Kaelin says. "Everyone will go to Kurdistan."

Kurdistan, an autonomous region about the size of Belgium in Iraq's north and east, was long off-limits to outsiders under the regime of Saddam Hussein. But today it welcomes foreign archaeologists and has relatively stable politics. "There has been no systematic research carried out here for the last 80 years," says Marta Luciani, an archaeologist at the University of Vienna who has long worked in Syria. Last year, her team mapped more than 50 previously unknown sites; one covers 5 hectares and is littered with pottery dating from 6000 B.C.E. to Islamic times.

Lost city. Bonacossi spent years digging here at Qatna in Syria but now cannot access his site or its finds.

Kurdistan's prehistory goes back even further. In the 1950s, archaeologists here found

*Eighth International Congress on the Archaeology of the Ancient Near East, 30 April–4 May, University of Warsaw. See http://www.8icaane.org.

CREDITS (TOP TO BOTTOM): F. D'AGOSTINO; PASQUALE SORRENTINO

Published by AAAS

40,000-year-old Neandertals buried in Shanidar Cave with what some argued were funeral rituals; the University of Cambridge in the United Kingdom is in talks to reopen the excavations, says university archaeologist Graeme Barker.

Kurdistan lies east of the oldest known Neolithic settlements, but researchers say sites in Kurdistan may offer clues to the transition to farming and settled life between about 10,000 B.C.E. and 7000 B.C.E. During initial excavations in the western Zagros Mountains in Kurdistan this March and April, for example, Roger Matthews of the University of Reading in the United Kingdom and colleagues found ancient grinding stones, hearths, architecture, and obsidian that may have come from as far as 200 kilometers away. Kurdistan is also the location of a main road linking Assyria with Babylon, two of the region's big powers in the first 3 millennia B.C.E. Harvard's Ur, who previously worked in Iran and Syria, hopes to probe two provincial Neo-Assyrian cities dating to about 800 B.C.E. and their hinterlands to understand how this great empire organized its cities, populations, and sophisticated waterworks.

Closed borders. Turmoil across the Middle East is making it tougher for foreign archaeologists to study early agriculture and urban development.

"Everyone is scrambling to get in on the ground floor" of work in Kurdistan, says Tina Greenfield, an archaeology Ph.D. candidate at the University of Cambridge. She is analyzing animal bones at ancient sites in Kurdistan but cannot cross the internal Iraqi border to the city of Mosul to study a bone collection there. But some archaeologists believe that the Kurdish promise is overblown. "It's the new hype," says Peter Pfälzner of the University of Tübingen in Germany, who worked in Syria.

Other alternatives for archaeologists are Persian Gulf states such as Oman and Kuwait (*Science*, 17 February, p. 790) and the Caucasus region of the former Soviet Union. Sandwiched between the Black and Caspian seas, the Caucasus has a bloody modern history that has kept it mostly off-limits, but Georgia and Azerbaijan now welcome digs.

In 2009, at a site called Kamiltepe in Azerbaijan, Barbara Helwing of the German Archaeological Institute in Berlin and her Azerbaijani colleagues found a large, round, brick platform dating to a surprisingly early 6000 B.C.E. The 24-meter-wide, 2.5-meter-high platform shows that "people are building monumental structures and transforming their lives" before archaeologists had expected such complexity, Helwing says. European teams are also starting work in neighboring Georgia, focusing on Neolithic connections among this region, the steppes to the north, and the emerging urban cultures of the south.

Other researchers are applying new tools to old samples. Heat, humidity, and poor bone preservation have made it extremely difficult to extract ancient DNA from Near Eastern bones, essentially leaving the discipline out of the genetic revolution. But at the meeting, biologist Henryk Witas of the University of Łódź in Poland presented preliminary evidence of ancient mitochondrial DNA from human teeth from a half-dozen skeletons at two sites in eastern Syria dated to various times in the 3rd millennium B.C.E. Most of the DNA was related to haplotype group M, which is not found in people living in the Middle East today but is common among those now living in northern Pakistan, India, and Tibet. Witas concluded that people migrated from the northern part of the Indian subcontinent along trade routes to the west as early as 2500 B.C.E.

This surprising conclusion was hotly disputed by others, who suspect that the M group once existed in the Near East but has been diluted since. "There is no archaeological evidence of Central Asian migration" before medieval times, notes archaeologist Maria Grazia Masetti-Rouault of the Sorbonne University in Paris, who excavated the Syrian sites. "It is way too premature to make any conclusions from this," adds Reinhard Bernbeck of the Free University in Berlin. But the paper demonstrated that, here as elsewhere, ancient DNA may serve as a key tool.

Archaeologists are also making heavier use of remote-sensing data. For example, Jesse Casana, an archaeologist at the University of Arkansas, Fayetteville, who once worked in Syria but now focuses on the United Arab Emirates, is experimenting with an instrument-heavy drone to help in mapping.

The closure of Syria is only the latest sign of trouble for ancient Near Eastern archaeology. Thanks to decades of lack of access to places like Iraq, prominent U.S. universities such as Harvard, Yale, and the University of Pennsylvania are not replacing retiring professors in the field. "Basically, we don't teach ancient Near Eastern archaeology anymore," says Yale's Hole. "It isn't viable." European researchers say they are experiencing a similar trend. "I'm one of the last Near Eastern archaeologists in Switzerland," adds Basel's Kaelin.

Others note that without sure places to dig, winning grants is difficult. "Once you lose excavation funding, it is hard to get it back," worries Udine's Bonacossi. Such uncertainties may create a vicious circle in which attracting new talent is tough. "There is a chilling effect," Ur says.

Still, the winds of change in the Middle East aren't all ill-favored for excavators. An American team and an Italian team briefly dug at sites in southern Iraq this winter and spring—the first time in nearly a decade that foreigners had returned to excavate in the heartland of ancient Sumer. During a brief survey last year, archaeologist Carrie Hritz of Pennsylvania State University, University Park, discovered that the ancient Mesopotamian city of Girsu is more than four times as large as researchers had thought; she plans a detailed survey this summer. Hritz and some colleagues of her generation are linking their efforts with the practical needs of Iraqis today, assisting in the search for clean water on the parched landscape.

Bonacossi remains bullish on the future of a field that has coped before with revolution, civil war, and arguably the world's most complex and explosive politics. "There are countries where we can still dig," he says. "The current trouble may just be history repeating itself." **–ANDREW LAWLER**

Published by AAAS

INFECTIOUS DISEASES

Can New Chemistry Make a Malaria Drug Plentiful and Cheap?

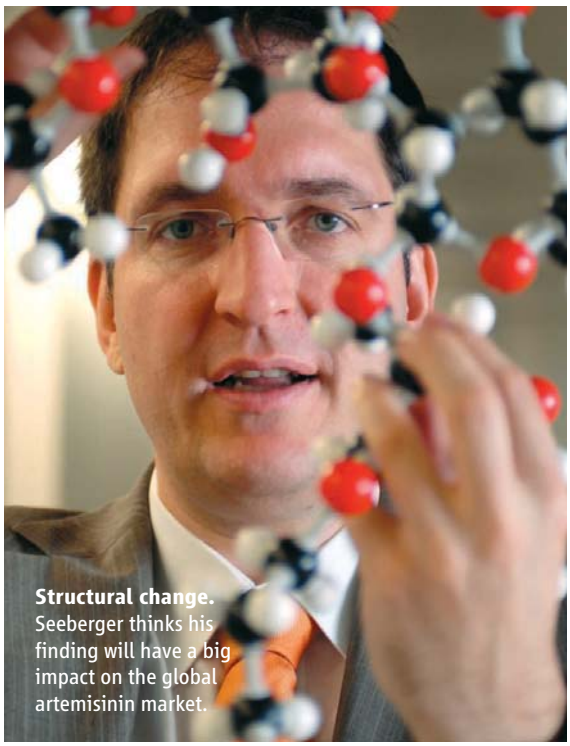
German chemist Peter Seeberger says he has developed a cheaper way to produce a key malaria drug. Now, he's trying to convince the rest of the world

HALLE, GERMANY—Television is not Peter Seeberger's natural medium. With several cameras trained on him, the 2-meter-tall chemist looks a little stiff in his dark suit, and he delivers his lines with an earnestness that is slightly at odds with the gee-whiz tone of the science show recorded here. He sounds like he's selling insurance.

But Seeberger, head of a team of 70 researchers at the Max Planck Institute of Colloids and Interfaces in Potsdam, came to sell an idea: that artemisinin, the world's most important antimalaria medication, can be produced much more cheaply and easily. In January, Seeberger published a paper outlining how a technique called flow chemistry might make a key step in the drug's production chain more efficient. If the promise comes true, it could be a boon for the global fight against malaria, because the current price, between $0.80 and $1.20 per treatment course, is still a major factor hampering access to artemisinin drugs.

Some scientists are impressed. "The impact of this is hard to overestimate," says Jack Newman, one of the founders of Amyris Inc. in Emeryville, California, a biotech company that has played a key role in other recent attempts to make artemisinin cheaper and more abundant. Prashant Yadav, an expert on malaria drug supply chains at the University of Michigan, Ann Arbor, says Seeberger's discovery is a "powerful technology, and it has the potential to increase global access to malaria medicine."

But Seeberger's method has yet to prove its mettle. It needs to be scaled up, and he can't say how much prices would come down if it worked. Using it in a large facility would require a massive investment, and so far, nobody has stepped up to the plate. What's more, pharma giant Sanofi will open a brand-new facility later this year to make artemisinin therapies based on Amyris's technology: yeast cells that produce a precursor of the drug. Although Seeberger says his discovery would complement that process, Sanofi says it's too late now to adopt it.



Structural change. Seeberger thinks his finding will have a big impact on the global artemisinin market.

CREDIT: ULRICH KLEINER

Plants to pills

Seeberger, 45, is known as a brilliant and driven researcher who's keen for his chemistry to make a difference—and who likes to try unconventional approaches. For most of his career, he has worked on sugars, or carbohydrates, which play many important roles in nature; they coat organs, cells, and molecules and facilitate all kinds of cellular communication.

When Seeberger started out as a scientist, carbohydrates were seen as career killers because they were hard to analyze and even harder to synthesize. Defying skepticism, he developed a new technique to string sugar molecules together and built a synthesizer that can churn out specific carbohydrate sequences on demand, much like peptide and nucleotide synthesizers. He has started a company to make the machine commercially available.

Since he came to the Max Planck Institute in 2008, Seeberger has used his expertise to develop carbohydrate-based vaccines against antibiotic-resistant *Staphylococcus aureus*, anthrax, plague, and malaria. The first animal studies of some candidate vaccines have been promising, although none has yet reached the stage of clinical studies. "But the artemisinin project is the most important discovery I have ever made in terms of immediate impact," he says.

To illustrate his work, he has brought a potted plant that looks a bit like oversized parsley with him to the TV studio. It's sweet wormwood (*Artemisia annua*), which has been used against fever for centuries in traditional Chinese medicine. In the 1970s, Chinese scientists isolated the active compound, artemisinin, and tested it as a malaria drug. It was the start of a worldwide revolution. Today, co-formulated with older drugs in so-called artemisinin-based combination therapies (ACTs), artemisinin and various close cousins have become the first line of defense against malaria.

But artemisinin is a difficult medicine to produce. Chemists can synthesize it from scratch, but that's prohibitively expensive. Instead, producers still rely on sweet wormwood, most of it grown in China, which makes artemisinin one of the last major drugs to be harvested from plants. *Artemisia* usually contains less than 1% artemisinin, however, and it takes almost 18 months from planting the seeds to extracting the compound. Supply has gone up and down, taking prices for drug companies on a roller-coaster ride from $400 to $1100 per kilogram.

Lower production costs would have a big impact on global health, Yadav says. In government-run clinics in most developing countries, malaria drugs are subsidized by big international donors. Cheaper drugs would mean they could help more patients. And many people in developing countries still rely on private pharmacies, where they often buy other, much cheaper and less effective drugs instead of ACTs. "There, every penny reduction really counts," Yadav says.

So far, the yeast cells developed by Amyris—which was co-founded by Jay Keasling of the University of California, Berkeley—seem to provide the best chance of a more stable artemisinin supply. Amyris, which is supported by the Bill and Melinda Gates Foundation, has licensed the technology to Sanofi, which has built a special plant in Garessio, Italy, where it plans to start producing ACTs later this year. But those cells

Published by AAAS

have one drawback: They don't produce the final product, artemisinin, but a precursor, artemisinic acid, that still needs to undergo a chemical reaction.

Artemisinin's activity depends on its so-called endoperoxide group, a bridge of two oxygen atoms that spans one of the molecule's three rings (see graphic, right). To make artemisinin out of artemisinic acid, chemists need to create the bridge by introducing reactive oxygen molecules called singlet oxygen. This can be done by shining light on "normal" oxygen molecules, a process known as photochemistry.

In a first step, artemisinic acid is reduced to dihydroartemisinic acid. This product is mixed with oxygen; the light activates the oxygen, which reacts with the acid to produce another precursor. Then, trifluoroacetic acid is added to the mix, cleaving a carbon ring in the molecule, which reacts with the molecular oxygen to produce artemisinin.

Sanofi plans to do this using so-called batch chemistry, in which a product is created via a series of steps in big vats, including, in this case, a specially developed glass reactor. But Seeberger says photochemistry is not well-suited to this classical approach because the bigger reaction vessels get, the less light they let in, and the less reactive oxygen is produced. Instead, he uses flow chemistry, a technique in which the reactions happen while the chemicals are flowing through a thin tube wrapped around a light source, which dramatically increases the volume in which reactive oxygen is produced.

"This approach makes a lot of technical sense," says Frank Gupton, a chemist at Virginia Commonwealth University in Richmond. Flow chemistry is ideally suited for reactions that need light, Gupton says, but chemical engineers have been slow to adopt it: "The chemical culture is really steeped in the batch chemistry mentality."

Waste not, want not?

But what will happen with Seeberger's discovery is still unclear. Sanofi's plant is about to open, and the company isn't going to bet on an entirely new technique that has yet to prove that it can be scaled up. In an e-mail to *Science*, the company calls Seeberger's solution "a clever approach," but says that "so far the competitivity of this technique has not been demonstrated."

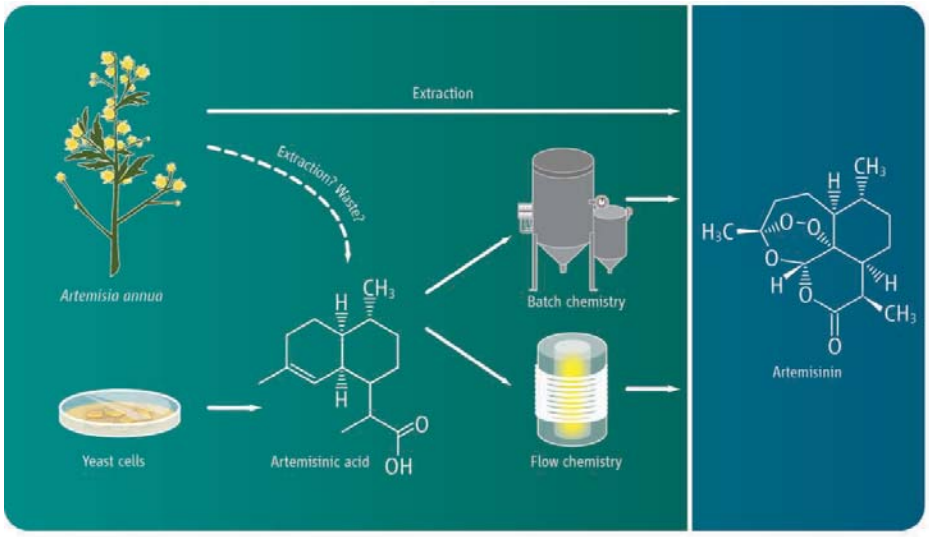
The ideal solution would be if other companies adopt the combination of Amyris's yeast cells and Seeberger's method, Yadav says; "then, the price for the drugs could go down significantly." But a spokesperson for OneWorld Health, the nonprofit pharmaceutical company that has backed Sanofi's project, says there are no plans to make the yeast cells available to any other party.

Seeberger says there are other ways to use his technique that don't rely on the production of artemisinic acid by yeast cells at all. Sweet wormwood plants often produce artemisinic acid in higher quantities than artemisinin, and Seeberger's initial idea was to use the artemisinic acid contained in waste, which extractors currently throw away. In a sample from extractors from Madagascar, Seeberger's team found 1% artemisinic acid. To his surprise, it also contained 5% dihydroartemisinic acid. The latter would be an even better starting material, eliminating the first reduction step.

But Yadav says the idea of turning waste into medicine is far-fetched. "It is a fairly fragmented market with multiple extractors. You would have to collect the waste from all these small companies and establish a supply system," he says.

Another option would be to ask extractors to start harvesting artemisinic or dihydroartemisinic acid directly from the plants, instead of artemisinin. Already, Seeberger is testing varieties of the plant to see how much of each compound they contain. But this would mean that the extraction companies have to make a major shift in their production process, requiring time and technical support from the global community.

On 19 April, Seeberger invited interested parties to a meeting in Berlin to explore the options. They included representatives of *Artemisia* growers and extractors, pharmaceutical companies GlaxoSmithKline and Boehringer Ingelheim, as well as the Clinton Foundation, UNITAID, and the German Agency for International Cooperation. (The Bill and Melinda Gates Foundation canceled at the last minute.) None of the funders wanted to discuss the meeting with *Science*. Seeberger says he was asked many critical questions—"But then the next day, my phone did not stop ringing." He is now in discussions with several interested parties, he says.



New route. Artemisinin is currently extracted from *Artemisia annua* plants. Later this year, Sanofi plans to start making the drug from genetically engineered yeast cells that produce the precursor artemisinic acid. Sanofi is using batch chemistry to convert artemisinic acid to artemisinin. Seeberger's flow chemistry process starts with artemisinic acid. The compound is also present in the plants, but this is not currently used.

But for the moment, he is soldiering on by himself. Since January, a new postdoc has taken over the project; by changing the solvent and the temperature and introducing an LED as the light source, he has increased the yield from 40% to 65%. And to start the scale-up, Seeberger has just ordered a new LED 10 times as big as the one he previously used. When it arrives, he wants to build a new reactor that can churn out a kilogram of artemisinin per day. The next reactor would be 10 times bigger again and capable of producing more than 3 tons of artemisinin per year, enough for a few million treatments.

Drumming up more interest was also one reason he agreed to be on the TV show. "So one day we might be able to say that this experiment has saved millions of lives?" the presenter of the TV show concludes as Seeberger's 5-minute appearance draws to a close.

"Yes, that is the hope," Seeberger answers.

–KAI KUPFERSCHMIDT

Kai Kupferschmidt is a science writer in Berlin.

CREDIT: G. GRULLÓN/*SCIENCE*

LETTERS

edited by Jennifer Sills

Forced Retirement Goes Out of Style

AAU President Hunter Rawlings III

IN THE INTERVIEW WITH J. MERVIS ("ON TEACHing, tuition, and talent," News Focus, 16 March, p. 1299), Association of American Universities (AAU) President Hunter Rawlings III puts forth the position that older faculty members should retire to make way for "the next generation to step forward." His advocacy of "forced" retirement is out of step with reality.

The socioeconomic model of retirement at 65 is no longer economically sustainable, and a substantial fraction of the "baby boom" generation have repeatedly indicated that they do not wish to retire. The scientific community should lead by example, demonstrating that people can and do make substantial contributions to their field and to society even at an advanced age.

The National Institutes of Health (NIH), AAU, and others should be concerned solely with funding the most promising projects. If there is a preponderance of support going to older researchers, it is hopefully because their projects appear to hold the most promise of significant achievement and not simply because of the past track record of those applicants. NIH might also consider the possibility that the increasing average age of grant recipients simply reflects the changing demographics of the United States in general.

Simply forcing older investigators to retire will not ensure that younger investigators of greater or at least equal talent will replace them. NIH, AAU, and their partners should be focused on locating and promoting the most talented scientists, regardless of age or academic standing.

ROY WILLIAMS

Department of Chemistry, University of Texas of the Permian Basin, Odessa, TX 79761, USA. E-mail: williams_r484@utpb.edu

Replication Initiative: Dangerous Logic

THE COMMENTS VOICED BY SOME SCIENTISTS IN response to Brian Nosek's project to systematically replicate studies in social psychology ("Psychology's bold initiative," S. Carpenter, News Focus, 30 March, p. 1558) are deeply worrisome. Reluctance to replicate data in a given field of science because it might harm the field's public image among other disciplines is symptomatic of a very dangerous logic. If such concerns are deemed valid, it seems natural to assume that scientists should also be worried about replicating their own labs' findings because of their standing among other laboratories, or that postdocs should not try to repeat experiments because it will make them look bad among their lab colleagues. The fact that this logic has spread among scientists is probably a large part of the reason why systematic replication studies have become so urgent in the first place. When the public image of a given field of science (or of science as a whole) becomes more important than establishing the truth of its findings in its value scale, one must conclude that something has gone awry along the way.

OLAVO B. AMARAL

Instituto de Bioquímica Médica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ 21941-902, Brazil. E-mail: olavo@bioqmed.ufrj.br

Replication Initiative: Prioritize Publication

THE OPEN SCIENCE COLLABORATION'S NEW platform for publishing the results of replicated experiments is commendable and indeed bold ("Psychology's bold initiative," S. Carpenter, News Focus, 30 March, p. 1558). However, as convenient and efficient as this platform may prove to be, it should not distract us from the need to revise and extend more conventional and reliable publication methods so as to include more replicated studies.

We must ensure that replicated studies, not just in psychology but across the sciences, are scrutinized just as closely as original studies, through the tried-and-tested peer-review system. To do this, we need better incentives and support for scientists who want to perform replications.

The argument that peer-reviewed journals should publish more failed replicated experiments is not new (*1*). A handful of journals publishing detailed methods and protocols have emerged. A broader solution could involve individual scientific disciplines launching their own peer-reviewed journal(s) devoted entirely to publishing the results of replicated experiments—both failed and successful. It would be important to ensure that these journals don't fall into the trap of a publication bias toward failed replications. Journals with this mission would help to raise the status of replication studies, and scientists would be less likely to see replications as a waste of their time and resources. In particular, postdoctoral and graduate-level scientists could benefit from this opportunity as a means of learning new experimental techniques and gaining early publications.

Systematic funding for replication studies should also be made available. Funding bodies such as the National Institutes of Health and National Science Foundation could designate awards for replication studies, with priority being given to the replication of novel experiments of international scientific importance.

Such initiatives are important, not just to

Letters to the Editor

Letters (~300 words) discuss material published in *Science* in the past 3 months or matters of general interest. Letters are not acknowledged upon receipt. Whether published in full or in part, Letters are subject to editing for clarity and space. Letters submitted, published, or posted elsewhere, in print or online, will be disqualified. To submit a Letter, go to www.submit2science.org.

address scientific fraud (*2*), but to ensure that reliability—one of the key principles upon which the authority of science is based—is publicly upheld. **SIMON NICHOLAS WILLIAMS**

Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA. E-mail: simonwilliams@northwestern.edu

References

1. J. Giles, *Nature* **442**, 344 (2006).
2. J. Crocker, L. M. Cooper, *Science* **334**, 1182 (2011).

Replication Initiative: Beware Misinterpretation

TWO NEW INITIATIVES ARE EXPLORING THE reproducibility of findings within psychology ("Psychology's bold initiative," S. Carpenter, News Focus, 30 March, p. 1558). One approach provides an online outlet for replication attempts, regardless of success or failure. The other approach evaluates reproducibility with new empirical replications. Proponents of these initiatives should be careful, because it is easy to misinterpret replication successes and failures in a field that uses statistics (*1*).

Psychologists establish empirical findings by using hypothesis tests that determine the probability (P) of data more extreme than observed under a null hypothesis. Equally important is power, which is the probability of rejecting a false null hypothesis. These probabilities reflect the inherent uncertainty in conclusions that can be drawn from a sampled set of data. Because of random sampling, even experiments that measure true positive effects will sometimes fail to reject the null hypothesis. For example, an experiment that just barely rejects the null hypothesis (e.g., $P = 0.05$) is expected to yield insignificant results almost half the time in future replications with the same sample size. A failure to replicate an established finding often ignites a debate about the differing conclusions, but in many cases there is no real conflict. Given the relatively low power of most psychological experiments, a failure to replicate should be common.

It is also possible to have too much replication success. Rejection of the null hypothesis at a rate inconsistent with the experimental power values indicates the possibility of publication bias (*2*, *3*). In such cases, the experimental findings are nonscientific because relevant data is missing.

The significance of replication studies should be assessed by carefully structured meta-analysis studies (*4*). **GREGORY FRANCIS**

Department of Psychological Sciences, Purdue University, West Lafayette, IN 47907, USA. E-mail: gfrancis@purdue.edu

References

1. V. Hedges, I. Olkin, *Statistical Methods for Meta-Analysis* (Psychology Press, New York, 1985).
2. J. P. A. Ioannidis, T. A. Trikalinos, *Clini. Trials* **4**, 245 (2007).
3. G. Francis, *Psychonomic Bull. Rev.* **19**, 151 (2012).
4. G. Cumming, *Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis* (Routledge, New York, 2012).

CORRECTIONS AND CLARIFICATIONS

Reports: "Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks" by T. P. Vogels *et al*. (16 December 2011, p. 1569). As written, the second sentence of the first paragraph on page 1570 implied that the authors were using a learning rule as observed by Woodin *et al*. in 2003 (cited as reference 17). In fact, they used a stereotypical, mathematically derived Hebbian learning rule with a negative offset (also in Woodin *et al*. 2003), in which coincident spike pairs potentiate the synapse (not observed in Woodin *et al*. 2003). References 18 and 19 were incorrect. Reference 18 should be J. S. Haas, T. Nowotny, H. D. Abarbanel, *J. Neurophysiol*. **96**, 3305 (2006). Reference 19 should be C. D. Holmgren, Y. Zilberter, *J. Neurosci*. **21**, 8270 (2001). The text that cites these references should read: "The strength of the inhibitory synapses was initially weak but could change according to a spike timing-dependent plasticity rule, in which near-coincident pre- and postsynaptic spikes induce potentiation of the synapse [(*18*), but see (*17*, *19*)]. Additionally, every presynaptic spike leads to synaptic depression (*17*) (Fig. 1C)." We thank F. Wörgötter, M. Tsodyks, and collaborators for discussions.

theBUZZ

The Motherhood Effect

On 2 March, a News & Analysis story titled "Is motherhood the biggest reason for academia's gender imbalance?" (J. Mervis, p. 1030) described a paper by two developmental psychologists who argue that we've been overlooking the obvious explanation for the drop-off of women in science. In a related News & Analysis story ("Half-time jobs, full-time scientists," 2 March, p. 1031), Mervis describes one couple's unusual solution to starting a family: splitting a tenure-track position. Many readers wrote in to comment on the two stories. Excerpts from some of those comments are below. You can read all the comments at http://comments.sciencemag.org/content/10.1126/science.335.6072.1030 and http://comments.sciencemag.org/content/10.1126/science.335.6072.1031.

A selection of your thoughts:

Motherhood is absolutely the reason I left research—after two postdocs. Now to get back in with a research grant is virtually impossible. Most "start-up grants" have a time limit from the awarding of a Ph.D. That is just the start of the many obstacles…. **—Evelyn Voura**

… The problem is not having a child, the problem is that men do not do 50% of the work when having a child. If men and women each did 50% of the work when having a child (as occurs in some relationships), then each would have the same career setback, and having a child would have no gender-specific effects. **—Adrian Liston**

…I think the study has missed the main reason why women may drop out of science, which is the same reason why many women reach less frequently the top positions at work: Paternity leave virtually does not exist…. If fathers can have paternity leave that is virtually as long as maternity leave, then there would not be any discrimination during hiring because male and female would bear the same "risk" of being less productive at work at a certain time of their careers. Fathers are not given the choice, which means that mothers are not either, in most cases. **—Veronique Le Roux**

These half-time faculty positions basically give in to the current terrible system of tenure-track jobs that are incompatible with having a family. Both men and women ought to be able to hold a properly paid full-time job that allows them to have children and spend time with their family and still have a realistic chance of winning tenure. I sympathize with couples who have used the half-time positions as a mechanism to cope with the current system, but the answer is not to encourage more half-time positions but to fix the system so these scientists can be properly rewarded for the full-time work they do. **—Michael Koelle**

SOCIOLOGY

Violence Tamed

Michael Hechter

A History of Violence contends that there has been an extraordinary secular decline in the incidence of violence in European history since the late Middle Ages. In the wake of two world wars, the Holocaust, Hiroshima and Nagasaki, to say nothing of other more recent horrors, this claim may seem counterintuitive. Once the per capita rate of violence is considered, however, it becomes easier to appreciate. Even so, violence appears in many different forms. Some think that the concept should include the threat faced by the prospect of nuclear annihilation during the Cold War. Others insist that the term is socially constructed, hence subject to different meanings in different times and places. Muchembled, a historian at the University of Paris XIII, avoids some of these difficulties by focusing primarily on homicide, for which abundant and relatively reliable measures exist in archival records.

Homicide is an act largely committed by adolescent and recently married males, from about the ages of 15 to 25. In medieval rural societies, where young males were compelled to wait for many years before acceding to the full rights accorded to married men, youthful violence was not only tolerated but sometimes even encouraged. Young males aspired to a macho ideal in order to prosper in a world that resembled Hobbes's state of nature far more than our own. Instead of taking out their social and sexual frustrations against their fathers and other authority figures, they turned against local peers and members of neighboring localities. The rate of violence declined precipitously in Europe, especially at the beginning of the 17th century and again in the 19th century. The number of fights between men using knives and swords fell, first among the aristocracy and then gradually and unevenly among the lower social classes. The use of violence to settle interpersonal disputes also decreased. In territories beyond the state's control, however, male aggression persisted in the forms of the duel and clan vengeance.

Muchembled seeks to determine the factors responsible for this overall fall in violence. Essentially he argues that the decline is due to a system of controls over male aggression. Since the aggressive tendencies of young males are universal, any stable society must find a way of satisfying them. Over time, male aggression was sublimated and put at the service of the modern state, especially during wartime. He speculates that the absence of war on European soil since 1945 has denied disaffected young men an opportunity to act out their aggressive impulses. As a consequence there has been an uptick of violence linked to mass sporting events, such as soccer riots, and gang conflict.

This system that developed to control violence has two parts: first, the deployment of sanctions levied initially by civic and then later by state authorities; second, the (largely unexplained) emergence of a social norm proscribing the use of violence to uphold honor, signal virility, and settle disputes. Whereas the supply of sanctions depended on measures taken by governing authorities, the antiviolence norm was promulgated by socializing institutions such as churches, schools, and armies. This norm is now so well ensconced that murder has become a virtual taboo in developed societies. Indeed, one of the main goals of basic training in the modern military is to rekindle young male recruits' repressed desire to shed blood (see Stanley Kubrick's film *Full Metal Jacket* for a graphic portrayal).

In contrast to claims by others, Muchembled argues that social control over male aggression preceded the rise of the state. Because the fortunes of towns depended on trade, civic authorities had an interest in preventing disorder as it discouraged commerce. If trade was responsible for the initial supply of sanctions, then population pressure, which increased all over Europe around 1520, had much to do with the demand for social control. Due to larger cohorts, young men had to wait longer than usual to attain their fathers' positions in the social structure. For their part, the adults deflected responsibility for punishing their children onto the authorities, who imposed respect for the law. In these conditions, the support of older generations increased the effectiveness of previously enacted laws against homicide and infanticide.

A History of Violence is broadly comparative within the European context and relies on a variety of evidentiary sources from lettres de rémission to modern detective novels. Muchembled doesn't address the causes of male aggression, but his story rests squarely on a foundation first laid in Sigmund Freud's *Civilization and Its Discontents* (*1*). This great essay holds that the renunciation of biologically determined aggressive impulses is the price that individuals must pay to live in an ordered society. Freud thereby anticipated one of the key behavioral assumptions used by today's evolutionary psychologists.

The book's claim that violence has decreased substantially in modern history is hardly novel. The German sociologist Norbert Elias first made the case in 1939 (*2*), arguing that prosocial norms promoting etiquette and discouraging violence diffused down the social structure from the nobility at Louis XIV's court at Versailles, to the gentry, and ultimately to the lower classes. Subsequently, books by Michel Foucault, Pieter

A History of Violence
From the End of the Middle Ages to the Present
by Robert Muchembled; Jean Birrell, Translator
Polity, Cambridge, 2011.
383 pp. S79.95, £60, €72.
ISBN 9780745647463.
Paper, $29.95, £19.99, €24.
ISBN 9780745647470.

The reviewer is at the School of Politics and Global Studies, Arizona State University, Tempe, AZ 85287–3902, USA.
E-mail: michael.hechter@asu.edu

CREDIT: SIPA PRESS

Spierenburg, and (most recently) Steven Pinker have supported Elias's conjectures about this long-term trend (*3*–*5*). For its theoretical sophistication, breadth of coverage of the relevant social science literature, and superior marshalling of the empirical evidence, Pinker's is the best of the bunch.

References

1. S. Freud, *Das Unbehagen in der Kultur* (Internationaler Psychoanalytischer, Vienna, 1930).
2. N. Elias, *Über den Prozess der Zivilisation: soziogenetische und psycogenetische Untersuchungen* (Falcon, Basel, 1939).
3. M. Foucault, *Surveiller et punir: naissance de la prison* (Gallimard, Paris, 1975).
4. P. Spierenburg, *A History of Murder: Personal Violence in Europe from the Middle Ages to the Present* (Polity, Cambridge, 2008).
5. S. Pinker, *The Better Angels of Our Nature: Why Violence Has Declined* (Viking, New York, 2011).

10.1126/science.1220392

SOCIOLOGY

Schooling Violence?

Claire L. Adida

In *Educations in Ethnic Violence*, Matthew Lange brings social scientific scrutiny to the dominant view that education means human, social, and political progress. In many cases, Lange argues, education can in fact be linked to greater ethnic violence. Such a contention is not novel. Social scientists have identified a positive correlation between education level and participation in violence in the United States (*1*) and beyond (*2*). Most recently, education has been causally linked to an increase in the perceived legitimacy of political violence among Kenyan schoolgirls (*3*).

The book's contribution is therefore not in the novelty of the argument. Instead, Lange (a sociologist at McGill University) pushes the debate forward with an analysis that consistently addresses the reader's objections. Indeed, his theoretical claims are controversial, and our instinct is to resist them. But he deploys his evidence in a rich empirical strategy that is both broad and deep, weaving together cross-national statistical analysis with a wide range of case studies and unveiling patterns that are consistent with the claim that education and violence are, in fact, positively correlated.

For his statistical tests, Lange draws on data from 1960 to 1999 across 160 countries. Controlling for factors typically associated with ethnic violence—such as a country's regime type; levels of economic development, ethnic diversity, and political discrimination; total and youth population sizes; and geographical characteristics that facilitate rebellion—he finds education significantly positively correlated with the level of ethnic violence. This quantitative analysis establishes a pattern that holds across time and space. But it does not tell a causal story. To do so, Lange turns to comparative historical analysis and explores the processes and mechanisms linking education and violence at a more micro level. In these case studies, educational achievement precedes ethnic violence; participants in ethnic violence (particularly violence entrepreneurs) are highly educated; and several mechanisms emerge that show how education affects the motivations and capacities of individuals to participate in violence. In both colonial and postcolonial Cyprus, for example, under the influence of Greece and Turkey, education "actively socialized Greek and Turkish Cypriots to hold strong and incompatible ethno-national identities." Furthermore, Greek Cypriot schools provided a highly disciplined student body with time on its hands to mobilize and staff the violent Greek Cypriot nationalist movement EOKA.

Educations in Ethnic Violence
Identity, Educational Bubbles, and Resource Mobilization
by Matthew Lange
Cambridge University Press, Cambridge, 2012. 254 pp. $90, £55. ISBN 9781107016293. Paper, $26.99, £16.99. ISBN 9781107602373.

All too aware that he is putting forth a counterintuitive argument, Lange seeks nuance in two important ways. First, he tries not to overreach with claims of causality. His quantitative analysis establishes statistical correlations and nothing more; that, he recognizes. Nonetheless, in his qualitative work he boldly speaks of a causal relationship between education and violence, ignoring the alternative explanation that both education and violence could be shaped by some third, omitted factor that would negate any causal effect of one on the other.

Second, Lange addresses the scope of his claim, delineating three conditions for his prediction to hold: ethnic divisions, resource scarcity, and ineffective political institutions. In his statistical analysis, he demonstrates that the positive relationship between education and violence holds best conditional on high ethnic diversity, low per capita gross domestic product, and poor political institutions. In his case studies, he covers a set of instances that are inconsistent with his predictions; he shows that in each of these, the scope conditions did not hold. For example, in the Indian state of Kerala, where educational attainment is high but ethnic violence low, the state government has been effective at promoting human development and the rule of law. Such political institutions mitigate the effect of education on violence. Similarly, the Quebec nationalist movement that has shaped Canada's political landscape since the 1960s never deteriorated into ethnic violence because the country benefits from both resource abundance and effective political institutions that limit the motivations of violence entrepreneurs and participants. Defining these scope conditions brings subtlety to a controversial argument. Yet it also takes us back to the problem of causality: If schools exercise no causal effect in Kerala, where political institutions are strong, how can we know for sure that they become causal in places where those institutions are weak? Might not those weak institutions be driving all the violence, relegating schools to an epiphenomenal role?

Lange recognizes this limitation and struggles with it throughout the book. Early on, he notes that "education is rarely an independent cause of ethnic violence but is usually a background condition that exerts its effect when combined with other factors." And he insists that "[w]hile suggesting that education is rarely a proximate or ultimate cause of ethnic violence, my conclusion in no way denies that education contributes to ethnic violence in influential ways." As a result, we are left with a compelling story about an important correlation, not a causal link. It is only in conjunction with other research dealing explicitly with identifying the causal effects of schooling (*3*) that we can gain more confidence in the causal nature of the relationship.

All the same, Lange makes a critical contribution with *Educations in Ethnic Violence*. The depth of explanation and breadth of evidence he brings to a counterintuitive, and uncomfortable, association leave us with a rather unsettling sense that not all good things go together.

References

1. A. Miller, L. Bolce, M. Halligan, *Am. Polit. Sci. Rev.* **71**, 964 (1977).
2. A. Krueger, J. Malečková, *J. Econ. Perspect.* **17** (4), 119 (2003).
3. W. Friedman, M. Kremer, E. Miguel, R. Thornton, "Education as liberation?" (National Bureau of Economic Research, NBER Working Paper No. 16939, Cambridge, MA, 2011); www.nber.org/papers/w16939.

10.1126/science.1220551

The reviewer is at the Department of Political Science, University of California, San Diego, La Jolla, CA 92093–0521, USA. E-mail: cadida@ucsd.edu

SOCIAL PSYCHOLOGY

Parochialism as a Central Challenge in Counterinsurgency

Current U.S. practice in Afghanistan may reify social divisions, which undermines institutions critical to postwar stability.

Nicholas Sambanis,[1*] Jonah Schulhofer-Wohl,[2] Moses Shayo[3]

America's power preponderance since the end of the Cold War has not translated into an ability to win quickly and decisively against insurgency. The U.S. military, designed to fight Soviet tanks on European battlefields, for the past decade has fought insurgents wearing flip-flops and using improvised explosives in Iraq and Afghanistan. Clear victories in counterinsurgency are rare, and these wars are costly (*1*) and long-lasting (table S1). Peace after civil wars, of which insurgencies are a subtype, is tenuous.

A clearing operation in Marja, Afghanistan.

PHOTO CREDIT: GORAN TOMASEVIC/REUTERS

Is self-sustaining peace an elusive goal for U.S. intervention? How can the conduct of counterinsurgency (COIN) be better designed to shift violent, fragmented societies to a peaceful equilibrium? We describe how scientific knowledge on the determinants and characteristics of human parochialism—the tendency to cooperate with and favor members of one's group—should change the way we approach these questions.

Insurgency and Parochialism

In an insurgency, an armed group or groups fight to depose the incumbent government by eroding its legitimacy and territorial control. Insurgency involves violence, but insurgents and civilian sympathizers also fill nonviolent roles, e.g., in intelligence, logistics, propaganda, service provision, and even governance. Throughout history, civilian support has been key to insurgents' ability to operate. Civilian support can be coerced but can also reflect commitment to the cause and prosocial behavior (e.g., hiding members of the resistance during the Nazi occupation of France; and providing food to American revolutionaries during the war of 1776).

Parochialism can be manifested in enhanced intragroup cooperation (e.g., by contributing to group-specific public goods or offering oneself to protect the group's security) but also in competitiveness and antisocial behavior toward rival outgroups (from diminished intergroup cooperation to preemptive strikes to neutralize threats from outgroups and attacks to increase the ingroup's resources and status) (*2*–*13*). COIN policies can make defeating insurgency harder by inadvertently activating ethnic or sectarian cleavages. If COIN reifies these social divisions, it can also undermine integrative institutions that cultivate a common national identity—institutions that are critical to the stability of postwar transitions.

Contemporary COIN tactics assume that civilians are uncommitted and seek to win their support by providing security and material incentives. COIN can be effective against opportunistic fighters and frightened civilians. However, the motive to fight and the impetus to support a particular side can vary, from marginalization (*14*); opportunistic profit-seeking (*14, 15*); fear, coercion, or revenge (*14, 16*, *17*); or political exclusion (*18*) to commitment to a group or cause (*19*). Evidence on the connection between violence and economic factors—such as employment and economic growth—is mixed (both across cases and, importantly, across regions or periods within particular cases) [see (*20*–*26*)]. Current approaches to COIN do not fully consider evidence on the determinants of parochial behavior in group settings. COIN operations can sow the seeds for future challenges to peace: As violence hardens group identities, counterinsurgency, which necessarily involves the use of force to secure territory, can strengthen the power of ethnic and/or local parochialism against efforts to gain the allegiance of the population.

The Logic and Practice of COIN

As it faced escalating violence in Iraq's civil war in 2006–07, the U.S. military endeavored to relearn the theory and practice of COIN. Current COIN doctrine, which reflects this experience, places primacy on a "population-centric" approach [(*27*) and supplementary materials (SM)]. Only by getting the population to side with the government can counterinsurgents achieve victory. This represents a shift from an "enemy-centric" approach, which centered on the pursuit and destruction of the insurgents [e.g., (*28*), SM].

"Clear-hold-build" operations, the cornerstone of COIN in Iraq and currently at work in Afghanistan, rest on two central assumptions: that the majority of civilians can be induced to support the government if their security is guaranteed and that insurgents and their civilian sympathizers can be "flipped" if given sufficient incentives (*29*, *30*) (SM). These assumptions derive from interpretation of a sample of historical cases of insurgency rather than scientific evidence on individual-level behavior in civil wars. The experiences of the British in Malaya, the French in Algeria, and the Americans in Vietnam have been particularly influential (*31*, *32*).

[1]Department of Political Science, Yale University, New Haven, CT 06520, USA. [2]Niehaus Center for Globalization and Governance, Woodrow Wilson School, Princeton University, Princeton, NJ 08544, USA. [3]Department of Economics, The Hebrew University of Jerusalem, Jerusalem 91905, Israel.

*Author for correspondence: nicholas.sambanis@yale.edu

Published by AAAS

The logic of clear-hold-build operations is as follows. Operations aim first to destroy insurgents' military capabilities through offensive action and to separate them from the population ("clear") so that civilians can feel secure. The emphasis can then shift to defensive military operations to protect the population from insurgents. Counterinsurgents demonstrate presence through patrols, assess and remedy the population's immediate needs, police the population to further separate out insurgents and deter reinfiltration, and target the insurgency's infrastructure ("hold"). Counterinsurgents then assist in improving economic and human development and help institute political reforms ("build"). These activities, aided by an information campaign to explain and justify the events taking place, set the stage for effective governance over the long-term.

Counterinsurgents adopt the view that only a small number of individuals are committed activists for the insurgents, while most of the population attempts to remain neutral and swings toward whichever side can better guarantee their safety. As security increases, material improvements can be used to ensure ongoing cooperation with the government. Coercion and provision of goods, both public and private, are counted upon to create sufficient incentives to support the government.

Understanding COIN Through Scientific Research on Individual Behavior

The COIN doctrine's characterization of individual behavior is incomplete. Security-seeking is a key motivation, but it is not always paramount, and human behavior is shaped by parochialism. Theoretical analysis (*2*) and archaeological and ethnographic evidence (*3*) suggest the emergence of parochial altruism in early humans where competition for resources favored groups with individuals willing to engage in conflict with outsiders on behalf of their group. Studies of group behavior further find that individuals do not automatically care about—or identify with—every group they belong to. Rather, identification is sensitive to factors shaping the salience of group boundaries, intra- and intergroup interactions, and group status. Some major empirical findings on parochialism in observed behavior are summarized in the table [for studies on stereotypes and prejudice, see (*33*)].

A first empirical finding is that even arbitrary assignment of individuals into groups is sufficient to trigger discriminatory behavior (*34*, *35*) and to generate altruistic behavior toward ingroup members and malevolent (envious) behavior toward outgroup members (*13*). Thus, any identity-based boundary that can plausibly define the government and counterinsurgent forces as belonging to an outgroup in reference to the target population can activate parochialism. This can hinder cooperation and counter the effectiveness of coercion and material incentives in gaining the population's support.

A second set of findings indicates that conditions, including institutions, which increase the salience of group membership or divisions, tend to increase parochialism (*5*, *12*, *35*). This is especially true with respect to intergroup conflict and violence. Individuals contribute more to their group under intergroup conflict when such contributions harm outgroup members (*9*). Even judges show significantly more bias against litigants from the opposite ethnic group after ethnically based terrorist activity near the court (*7*). There is also evidence that indiscriminate violence and "collateral damage" polarize the population (*17*, *36*). Coercion and violence directed by counterinsurgents against the local population are therefore uniquely problematic because they harden group boundaries.

This is an important lesson. Consider evidence on civil wars since 1945. Over half are ethnic (*37*), and ethnicity figures prominently in historical case studies of even wars that are not coded as being of an "ethnic" nature. Yet most empirical studies find no association between ethno-linguistic fractionalization and civil war onset (*38*). Others find a positive correlation between ethnic polarization (few large groups rather than many small ones) and conflict. But even then, polarization by itself explains very little of the variation in the incidence of civil war. Countries that are similar in ethnic structure—as well as along geographic, economic, and political lines—still exhibit quite different levels of conflict (*39*). Some highly diverse countries—like Angola, Indonesia, and Sudan—have experienced periods of intense violent conflict, whereas similarly diverse countries—like Tanzania, Zambia, and Brazil—

Behavioral evidence on parochialism

Setting	Data source*	Task	Main findings†	Factors shown to enhance parochialism
Dictator games	Lab (*13*), Field (*8*, *44*, *11*)	Allocate endowment between self and other.	Ingroup bias	Mere categorization into groups (*13*); subjective closeness to one's ethnic group (*8*); third-party punishment by ingroup member (*44*); mutual knowledge of coethnicity (*11*).
Minimal group paradigm	Lab	Allocate resources between anonymous ingroup and outgroup members	Ingroup bias	Mere categorization into groups (*34*, *35*); high group status (*42*).
Public goods games	Lab (*12*, *9*, *4*), Field (*5*, *10*, *11*)	Allocate endowment between self and contribution to group. Zero contribution maximizes own payoff; full contribution maximizes total (ingroup) payoffs. In some games contributions also affect outgroup payoffs.	Higher cooperation with ingroup than with outgroup members	Random assignment to platoons (*10*); Intragroup interaction (*12*); intergroup competition (*12*, *9*); segregated institutions (*5*).
Voting games	Lab	Vote over redistribution of income.	Ingroup bias	Low monetary cost for supporting group (*40*).
Judical decisions in court	Natural	Award or deny monetary transfers between litigants in civil cases.	Ethnic ingroup bias	Recent ethnic violence in vicinity of court (*7*).
Time and risk preference elicitation	Lab	Choose between receiving money earlier and receiving a larger amount later; choose between a sure sum and a lottery.	Conformity to ingroup norms	Salience of group membership (*45*).

*Lab: lab experiments; field: field experiments; natural: naturally occurring data. †Ingroup bias: preferential treatment of members of one's group.

Published by AAAS

have been relatively peaceful. It is not ethnicity per se but ethnicity made salient by violent conflict that leads to a vicious cycle in which violence and parochialism reinforce each other (*39*). COIN operations can inadvertently fuel this cycle.

Third, actively supporting one's group is sensitive to the cost of doing so (*40*). Parochialism does not trump other motives under all conditions. On the one hand, this is consistent with clear-hold-build. On the other hand, group loyalty is often cultivated in war, especially among combatants, and is widely considered a primary motivation for risking one's life in battle (*41*). During the American Civil War, fewer than 10% of Union soldiers deserted—although chances of being caught were low and risk of death if arrested insignificant—and consistent with experimental results, desertion was lower in more homogeneous units (*6*). Moreover, the government cannot maintain the high level of coercive force used during COIN over the long run, particularly when its efforts are supported by foreign forces. Parochialism could manifest itself in the postwar period, hindering the transition to peace. And material incentives alone are unlikely to be sufficient to guarantee the support of a population where ethnic or sectarian identification is entrenched and the government is part of the outgroup. Indeed, because parochialism tends to increase with the status of the relevant ingroup (*42*), material inducements to key social groups might even be counterproductive. The high risk of civil war recurrence and the difficulty in reaching and implementing agreements for power-sharing and integration of the rebels into the national army (table S1) are also indicative of these problems.

In Afghanistan, a large randomized control trial finds that development aid improved perceived security and attitudes toward the government only in districts with already low levels of violence (*23*). Where the level of conflict was high, aid could not buy loyalty or improve security. Ethnicity alone does not explain this. Residents of the most violent Pashtun regions, but not other Pashtun regions, maintained negative attitudes toward the government, despite welfare improvements. A survey experiment in the Pashtun-dominated southern provinces also questions whether aid had any effect on attitudes (*24*). Overall, the effects of COIN-related strategies are far from uniform. Development aid has exacerbated violence in some cases (*25*) and improved security in others (*21*), and the effect depends on the time period or region analyzed and the nature of the aid programs and types of violence considered (*22*). Existing evidence does not allow us to speculate as to how much the activation of parochialism accounts for variation in COIN success. It is clear, however, that short-term, strategic alliances across group boundaries are no indicator of parochialism's absence, although sometimes interpreted as such. Pragmatic alliances forged during conflict rarely reflect deep convictions. They are also no guarantee that, absent war's pressures, parochialism will not shape postconflict cooperation patterns.

Scientific research has yet to establish the relative importance of parochialism in explaining human behavior compared with material incentives. There is also need for more experimental evidence on the antisocial aspects of parochialism and the long-term effects of conflict on parochialism. Yet existing evidence is sufficient to raise concerns about the potential for intergroup cooperation in societies where COIN-related violence has reified ethnic, religious, or other cleavages.

The current approach to COIN is supposed to have worked in Iraq. This seemingly validates assumptions about individual behavior in insurgencies. Yet U.S. policy-makers may have drawn the wrong lessons. The 2007 military "surge" in Iraq did result in security gains (*21*), but clear-hold-build cannot claim all—or even most—of the credit for COIN's apparent success. Rather, the Sunni "awakening" was key to the defeat of al-Qaida in Iraq (AQI) and led to the transformation of the Iraqi army into a national, integrative institution in the making. The threat of Iranian domination and Sunni resentment of AQI's indiscriminate violence pushed Sunni elites to collaborate with the United States and the Shia-led government (*43*). The Iraq experience does not travel to Afghanistan, where COIN is not assisted by a common external threat or a preexisting strong national identity that can bring the insurgent factions together to side with the government. In the presence of social divisions, COIN as currently practiced can both compound the difficulty of defeating the armed challenge and make nation-building even harder.

References and Notes

1. A. Belasco, The cost of Iraq, Afghanistan, and other global war on terror operations since 9/11 (Congressional Research Service Report RL 33110, CRS, Washington, DC, 2011).
2. J.-K. Choi, S. Bowles, *Science* **318**, 636 (2007).
3. S. Bowles, *Science* **324**, 1293 (2009).
4. C. K. De Dreu *et al.*, *Science* **328**, 1408 (2010).
5. M. Alexander, F. Christia, *Science* **334**, 1392 (2011).
6. D. Costa, M. Kahn, *Q. J. Econ.* **118**, 519 (2003).
7. M. Shayo, A. Zussman, *Q. J. Econ.* **126**, 1447 (2011).
8. C. M. Fong, E. F. P. Luttmer, *Am. Econ. J. App. Econ.* **1**, 64 (2009).
9. G. Bornstein, *Pers. Soc. Psychol. Rev.* **7**, 129 (2003).
10. L. Goette *et al.*, *Am. Econ. Rev.* **96**, 212 (2006).
11. J. Habyarimana, M. Humphreys, D. Posner, J. Weinstein, *Am. Polit. Sci. Rev.* **101**, 709 (2007).
12. C. Eckel, P. J. Grossman, *J. Econ. Behav. Organ.* **58**, 371 (2005).
13. Y. Chen, S. X. Li, *Am. Econ. Rev.* **99**, 431 (2009).
14. M. Humphreys, J. Weinstein, *Am. J. Pol. Sci.* **52**, 436 (2008).
15. F. Christia, M. Semple, *Foreign Aff.* **88**, 34 (2009).
16. C. Blattman, *Am. Polit. Sci. Rev.* **103**, 231 (2009).
17. L. Condra, J. H. Felter, R. K. Iyengar, J. N. Shapiro, The effect of civilian casualties in Afghanistan and Iraq (NBER Working Paper 16152, National Bureau of Economic Research, Cambridge, MA, 2010).
18. A. Abadie, *Am. Econ. Rev. Papers Proc.* **96**, 50 (2006).
19. E. J. Wood, *Insurgent Collective Action and Civil War in El Salvador* (Cambridge Univ. Press, New York, 2003).
20. E. Berman *et al.*, *J. Conflict Resolut.* **55**, 496 (2011).
21. E. Berman *et al.*, *J. Polit. Econ.* **119**, 766 (2011).
22. R. Iyengar, J. Monten, M. Hanson, Building peace: The impact of aid on the labor market for insurgents (NBER Working Paper 17297, NBER, Cambridge, MA, 2011).
23. A. Beath, F. Christia, R. Enikolopov, Winning hearts and minds through development: Evidence from a field experiment in Afghanistan (Political Science Working Paper No. 2011-14, MIT, Cambridge, MA, 2012).
24. J. Lyall, K. Imai, G. Blair, Explaining support for combatants during wartime: A survey experiment in Afghanistan. (Simons Paper No. 17, Simon Fraser Univ., Vancouver, BC, 2011).
25. B. Crost, J. H. Felter, P. Johnston, Aid under fire: Development projects and civil conflict (Working Paper, Department of Economics, Univ. of Colorado, Denver, 2011); http://econ.ucdenver.edu/bcrost/research/ .
26. A. Krueger, J. Maleckova, *J. Econ. Perspect.* **17**, 119 (2003).
27. U.S. Department of the Army, *Counterinsurgency* (U.S. Army Field Manual No. 3-24, U.S. Department of the Army, Washington, DC, 2006).
28. N. Leites, C. Wolf Jr., *Rebellion and Authority: An Analytic Essay on Insurgent Conflicts* (Markham, Chicago, 1970).
29. D. J. Kilcullen, *Counterinsurgency* (Oxford Univ. Press, Oxford, 2010).
30. U.S. Department of the Army, *Commander's Guide to Money as a Weapons System* (Handbook No. 09-27, Center for Army Lessons Learned, Leavenworth, KS, 2009).
31. D. Galula, *Counterinsurgency Warfare: Theory and Practice* (Praeger, New York, 1964).
32. J. A. Nagl, *Counterinsurgency Lessons from Malaya and Vietnam: Learning to Eat Soup with a Knife* (Praeger, Westport, CT, 2002).
33. J. Dovidio, S. Gaertner, in *Handbook of Social Psychology*, S. T. Fiske, D. Gilbert, G. Lindzey, Eds. (Wiley, New York, ed. 5, 2010), vol. 2, pp. 1084–1121.
34. H. Tajfel *et al.*, *Eur. J. Soc. Psychol.* **1**, 149 (1971).
35. R. Y. Bourhis, A. Gagnon, in *Blackwell Handbook in Social Psychology: Intergroup Processes*, R. Brown, S. Gaertner, Eds. (Blackwell, Oxford, 2001), vol. 4, pp. 89–111.
36. D. A. Jaeger *et al.*, *J. Public Econ.* **96**, 354 (2012).
37. M. W. Doyle, N. Sambanis, *Making War and Building Peace: United Nations Peace Operations* (Princeton Univ. Press, Princeton, NJ, 2006).
38. H. Hegre, N. Sambanis, *J. Conflict Resolut.* **50**, 508 (2006).
39. N. Sambanis, M. Shayo, Social identification and ethnic conflict (Leitner Working Paper Series, Yale Univ., New Haven, CT, 2012); http://ssrn.com/abstract=1955111.
40. E. F. Klor, M. Shayo, *J. Public Econ.* **94**, 269 (2010).
41. S. A. Stouffer *et al.*, *The American Soldier: Combat and Its Aftermath* (Princeton Univ. Press, Princeton, NJ, 1949), vol. II.
42. B. A. Bettencourt *et al.*, *Psychol. Bull.* **127**, 520 (2001).
43. S. Biddle, *Perspect. Polit.* **6**, 347 (2008).
44. H. Bernhard *et al.*, *Nature* **442**, 912 (2006).
45. D. J. Benjamin *et al.*, *Am. Econ. Rev.* **100**, 1913 (2010).

Supplementary Materials

www.sciencemag.org/cgi/content/full/336/6083/805/DC1

10.1126/science.1222304

Published by AAAS

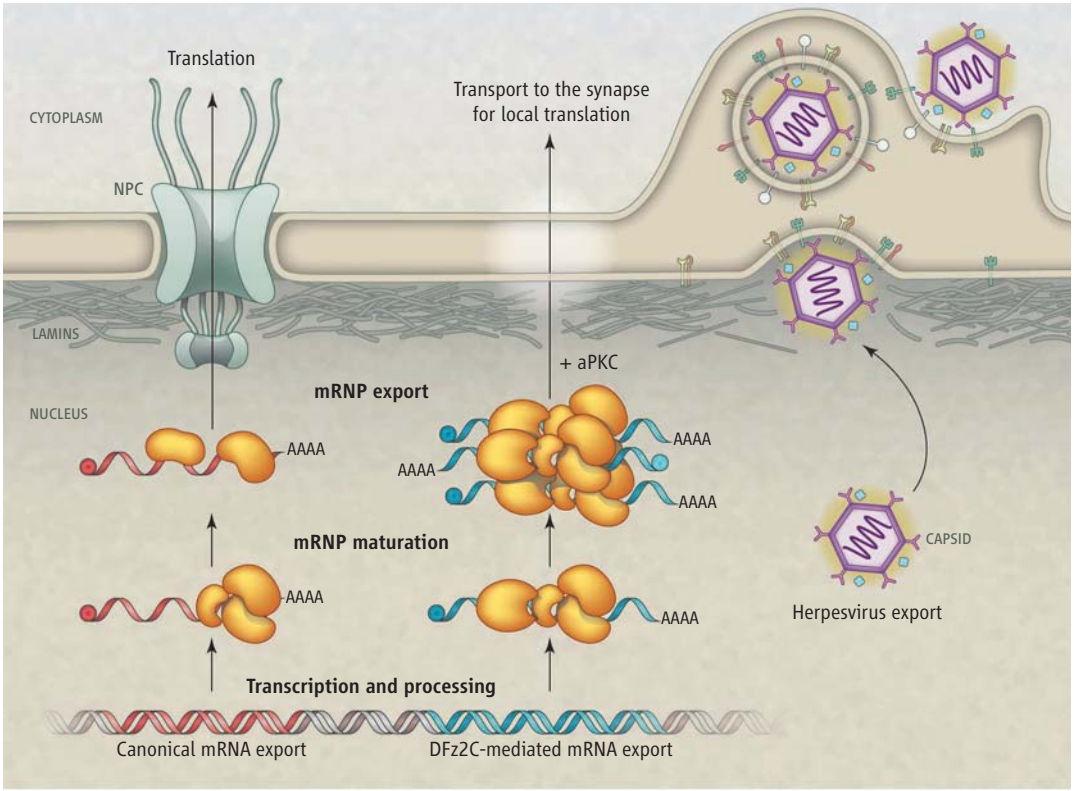
CELL BIOLOGY

An Alternative Route for Nuclear mRNP Export by Membrane Budding

Nuclear export of mRNA, previously thought to happen exclusively through nuclear pore complexes, may also occur via a membrane-budding mechanism.

Ben Montpetit and Karsten Weis

In eukaryotes, the double membrane of the nuclear envelope protects and organizes the genetic information that is stored as DNA in the nucleus. To use this information, the DNA must be transcribed into mRNA, packaged with proteins into a messenger ribonucleoprotein (mRNP) complex, and transported to the cytoplasm, where the information contained within the mRNA is translated into the corresponding protein. Until now, nuclear export of mRNP particles was thought to occur solely through nuclear pore complexes (NPCs)—transport channels inserted into the nuclear envelope at fusion sites between the inner and outer nuclear membranes (*1*). However, new findings by Speese *et al*. (*2*) suggest an alternative and unexpected route for a cellular mRNP to exit the nucleus, by membrane budding.



A new route. mRNP particles exit the nucleus via the NPC (**left**) or via a membrane-budding mechanism involving lamin C and atypical PKC (aPKC) that could enable export of larger translation-competent mRNPs or allow co-regulation of multiple mRNPs (**center**). This noncanonical export pathway appears similar to the nuclear egress of herpes-type viruses (**right**). Diagram not to scale.

NPCs create highly selective conduits through which cellular material can pass between the nucleus and cytoplasm (*1*). Nuclear transport receptors couple transport to energy consumption to drive differential localization of cargo against steep concentration gradients. Controlling the nuclear and cytoplasmic localization of interacting molecules provides a potent means of regulating diverse cellular processes. An example of this is the nuclear import of *Drosophila* Frizzled-2 (DFz2), the receptor of the secreted signaling protein Wingless (Wg) at neuromuscular junctions. Upon binding by Wg, DFz2 is endocytosed in the postsynaptic cell, cleaved, and its carboxyl terminus (DFz2C) is imported into the nucleus where it plays an important, but poorly understood role in synapse differentiation (*3*, *4*).

To better characterize the nuclear function of DFz2C, Speese *et al*. followed up on the observation that DFz2C localizes to discrete foci in the nucleus (*3*, *5*). Using a combination of imaging techniques, they discovered that DFz2C foci are closely associated with lamin C at the nuclear periphery and contain mRNAs coding for postsynaptic proteins. Intriguingly, the authors provide evidence that the DFz2C foci facilitate the nuclear exit of these mRNAs by a nuclear membrane budding mechanism. Formation of the DFz2C foci requires both lamin C and protein kinase C (PKC), and foci disruption severely affects synapse morphology and function. This egress is mechanistically similar to a pathway used by herpes-type viruses to liberate assembled viral capsids from the host nucleus (see the figure), which relies on PKC activity and lamin phosphorylation to locally remodel the nuclear envelope (*6*–*8*).

Many mechanistic questions remain, and as a first step it will be critical to block NPC function to test whether this export pathway is truly NPC-independent and to strengthen the conclusion that mRNPs traverse the nuclear envelope via membrane budding. Fortunately, single-molecule imaging techniques have been recently developed that should allow the visualization of this noncanonical mRNP export event and discern the relationship, if any, with NPC-mediated export (*9*, *10*). Viruses often take advantage of preexisting host pathways for their replication, and the parallel between the nuclear exit of herpesvirus capsids and that of DFz2C-mRNPs suggests that membrane budding is a general cellular mechanism for export of mRNPs and other cargoes, which will now be important to test. Future stud-

Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720–3200, USA. E-mail: bmontpet@berkeley.edu; kweis@berkeley.edu

Published by AAAS

ies are also required to identify the molecular machinery involved in both the physical and regulatory aspects of nuclear budding to fully characterize this fascinating new export pathway.

NPC-mediated transport is highly efficient, so why is there a need for an alternate export pathway? One possibility is that some cargoes are too large for export through an NPC. Alternatively, packaging multiple transcripts in a single large export particle may provide the means to cotransport, colocalize, and/or cotranslate functionally related mRNAs. Another possibility may be in cases where mRNP complexes need to be transported in a translationally repressed state, requiring maintenance of the mRNP structure, which can be disrupted during regular export of large mRNPs (*11*). Finally, a second export pathway could be subjected to alternative regulation, enabling the expression of specific mRNAs under conditions where general export is inhibited, thus allowing a new layer of regulation of the cellular gene expression program. Given these far-reaching implications, it will be critically important to explore all facets of this novel mRNA export process.

References

1. J. D. Aitchison, M. P. Rout, *Genetics* **190**, 855 (2012).
2. S. D. Speese *et al.*, *Cell* **149**, 832 (2012).
3. D. Mathew *et al.*, *Science* **310**, 1344 (2005).
4. B. Ataman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 7841 (2006).
5. B. Ataman *et al.*, *Neuron* **57**, 705 (2008).
6. T. C. Mettenleiter *et al.*, *Curr. Opin. Microbiol.* **9**, 423 (2006).
7. W. Muranyi *et al.*, *Science* **297**, 854 (2002).
8. R. Park, J. D. Baines, *J. Virol.* **80**, 494 (2006).
9. A. Mor *et al.*, *Nat. Cell Biol.* **12**, 543 (2010).
10. D. Grünwald *et al.*, *Nature* **475**, 333 (2011).
11. H. Mehlin *et al.*, *Cell* **69**, 605 (1992).

10.1126/science.1222243

CHEMISTRY

Active Site of an Industrial Catalyst

Surface steps caused by stacking faults or twin boundaries in copper nanoparticles are key to the activity of the methanol synthesis catalyst.

Jeffrey P. Greeley

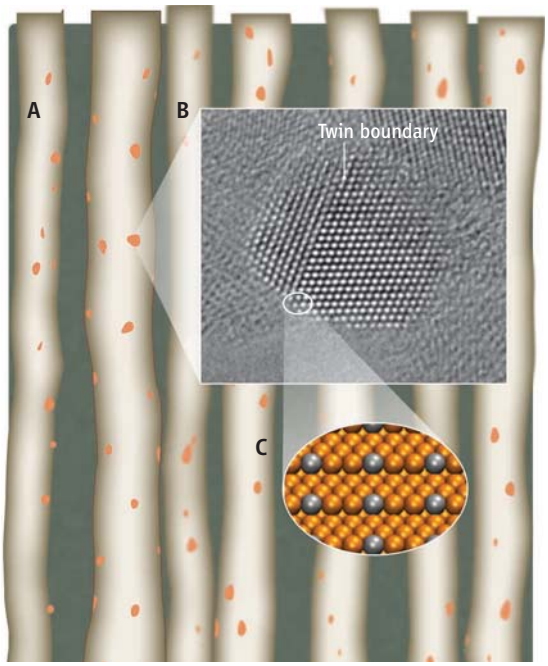
The active site of a heterogeneous catalyst can be thought of as the ensemble of atoms that directly catalyzes a reaction. Knowledge of the structure and composition of the active site is crucial for understanding and improving the properties of catalysts (*1*, *2*). The advent of atomically resolved in situ spectroscopies, enhanced surface science techniques, and accurate electronic structure calculations has led to important advances in understanding single crystal–based catalysts and metal nanoparticles (*3*–*7*). However, little progress has been made in probing the active sites of more challenging catalysts, in which interactions between the active metals and the oxide supports are highly synergistic and sensitive to the environmental and reaction conditions (*8*). On page 893 of this issue, Behrens *et al*. (*9*) report a comprehensive experimental and theoretical analysis of the active site structure of one such catalyst that is of crucial importance in industry.

The material studied by the authors is the methanol synthesis catalyst, which is composed of a mixture of copper and zinc oxide on an alumina support. This catalyst has been heavily optimized in industry, but the basis of its activity remains poorly understood (*10*). Previous efforts to understand its properties have relied on model systems to determine fundamental catalytic properties. However, these model studies have generally failed to capture important qualitative features of the methanol synthesis catalyst, because the strong metal-support interaction is highly sensitive to the catalyst structure and environmental conditions.

To overcome these limitations, Behrens *et al*. synthesized and analyzed a series of materials with structures and properties closely analogous to those of the industrially optimized catalysts. By introducing controlled variations into the synthesis and comparing the resulting properties to those of the industrial catalyst, they were able to systematically isolate important structural features that determine the catalytic performance.

A key finding is that the density of stacking faults or twin boundaries in the catalysts' copper nanoparticles, as determined by neutron diffraction experiments, correlates linearly with the measured intrinsic catalytic activity per surface area. Using high-resolution transition electron microscopy (TEM), the authors also show that steps or kinks are formed at the intersections between stacking faults or twin boundaries and copper nanoparticle surfaces (see the figure, panel B). These observations in turn suggest that steps constitute an important component of the active sites for methanol synthesis catalysts. To further verify this hypothesis, the authors carried out extensive periodic density functional theory (DFT) calculations of methanol synthesis pathways on both flat and stepped copper surfaces. They found a systematic lowering of all relevant reaction barriers on the step edges, confirming that these defects will show higher intrinsic activity for methanol synthesis from either carbon monoxide or carbon dioxide.

Although these data clearly show the importance of copper defects in catalyzing the methanol synthesis reaction, the promising properties of these steps alone



The importance of steps. Model studies have failed to capture important features of the industrially important methanol synthesis catalyst (**A**). Behrens *et al*. now show that stacking faults or twin boundaries are likely to play a key role in this activity by causing steps to be formed on the copper nanoparticle surfaces (**B**). Zinc atoms are drawn to these steps (**C**), further enhancing catalytic activity.

Center for Nanoscale Materials, Argonne National Laboratory, Argonne, IL 60439, USA. E-mail: jgreeley@anl.gov

Published by AAAS

cannot fully account for the performance of industrial methanol synthesis materials. Copper nanoparticles by themselves show much lower catalytic activities than the industrial catalysts; to achieve better performance, it is necessary to add zinc oxide. Using a combination of x-ray photoelectron spectroscopy and TEM, the authors found that zinc atoms can be drawn to the copper surface by partial reduction via one of the reactants, dihydrogen; this segregation is reversible upon high-temperature oxidation of the catalyst.

The tunability of the surface zinc concentration under reaction conditions suggests that this species may also play an important role in catalyzing methanol synthesis. Additional DFT calculations confirmed that zinc atoms in the copper steps (see the figure, panel C) further decrease key reaction barriers, thereby enhancing catalytic activity. The authors further postulate that zinc oxide has nearly optimal reducibility for catalyzing the methanol synthesis reaction; it is sufficiently reducible to bind to oxygenated reaction intermediates under industrially relevant conditions, but not so reducible that an undesirable copper-zinc bulk alloy forms.

Behrens *et al.*'s determination of the active site of the industrially important methanol synthesis catalyst solves a long-standing puzzle in heterogeneous catalysis. Now that the key structural features have been identified, it may be possible to effect further improvements in the catalyst. For example, it remains to be shown whether practical strategies exist for further increasing the step/defect density in copper nanoparticles. Additionally, the hypothesis that zinc oxide has ideal reducibility to catalyze methanol synthesis could be further evaluated if a method of quantitatively relating the reducibility to the catalytic properties could be found. In a more general sense, the approach followed in this work should serve as an excellent model for the study of active sites in other catalytic materials where analysis of simplified model systems is not sufficient because of the strong coupling of metals, supports, and the reactive environment.

References

1. G. A. Somorjai, *Introduction to Surface Chemistry and Catalysis* (Wiley, New York, 1994).
2. J. Greeley, J. K. Nørskov, M. Mavrikakis, *Annu. Rev. Phys. Chem.* **53**, 319 (2002).
3. K. Honkala *et al.*, *Science* **307**, 555 (2005).
4. M. Neurock, *J. Catal.* **216**, 73 (2003).
5. A. Cuesta, *J. Am. Chem. Soc.* **128**, 13332 (2006).
6. S. Dahl *et al.*, *Phys. Rev. Lett.* **83**, 1814 (1999).
7. P. Ferrin, M. Mavrikakis, *J. Am. Chem. Soc.* **131**, 14381 (2009).
8. P. L. Hansen *et al.*, *Science* **295**, 2053 (2002).
9. M. Behrens *et al.*, *Science* **336**, 893 (2012); 10.1126/science.1219831.
10. J. B. Hansen, P. E. Højlund Nielsen, in *Handbook of Heterogeneous Catalysis,* G. Ertl, H. Knözinger, F. Schüth, J. Weitkamp, Eds. (Wiley-VCH, Weinheim, Germany, 2008), pp. 2920–2949.

10.1126/science.1222974

CHEMISTRY

NMR Tools for Determining the Structure of Plutonium Materials

The ability to observe plutonium-239 magnetic resonance signals should aid in environmental studies and the development of waste-storage materials.

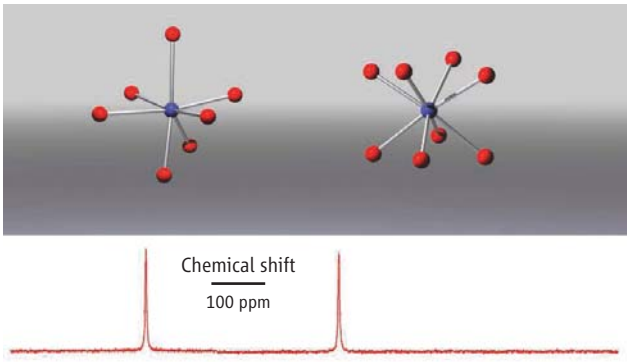
Thomas E. Albrecht-Schmitt

Nuclear magnetic resonance (NMR) is an extremely powerful method for determining molecular structures in a rapid and nondestructive way. The technique is typically applied to nuclei that have a spin quantum number of ½, which fortunately includes many of the nuclei in organic molecules, such as ^{1}H, ^{13}C, and ^{31}P. NMR spectroscopy yields different signals for specific nuclei—"chemical shifts"—because different local arrangements of nearby nuclei and electronic bonds perturb the local magnetic field of a nucleus. The patterns of chemical shifts can be determined for compounds of known structure, and structures of unknown compounds can often be ascertained on the basis of the number and location of signals in the spectrum; their relative intensity of signal reflects the number of nuclei experiencing that specific chemical environment. The last of the heavy nuclei with $I = ½$ for which the NMR spectrum was undetermined has been ^{239}Pu, despite decades of effort, but on page 901 of this issue, Yasuoka *et al.* (*1*) report a technique for ^{239}Pu NMR.

The problem of measuring a ^{239}Pu NMR signal has not been limited by the abundance of NMR-active nuclei in samples—the weapons-grade plutonium that has been synthesized by nuclear-enabled countries is typically ~94% ^{239}Pu. (By contrast, the ^{17}O isotope has a very low natural abundance of 0.04%, and deliberately enriched compounds must be synthesized to obtain useful spectra.)

An intrinsic problem in obtaining ^{239}Pu NMR spectra is that plutonium is probably the most complex element in the periodic table in terms of its oxidation chemistry—it can equilibrate four different oxidation states in solution. No other element can do this, and this fluxional behavior makes characterizing plutonium samples difficult, particularly solids. Also, the most prevalent oxidation state of plutonium under many conditions is Pu(IV). Plutonium in this oxidation state has a $5f^4$ electron configuration and typically possesses a nonmagnetic ground state. Thus, many common techniques for determining the electronic properties of metal ions, such as magnetic susceptibility and electron spin resonance, yield little useful information, making it difficult to assign spectra on the basis of known structural data.

Beyond the intrinsic nature of the samples, there are three aspects of ^{239}Pu NMR that



Signal efforts. Simulated spectra are shown for two different Pu environments in a complex material with a low-symmetry seven-coordinate site (**left**) and a high-symmetry eight-coordinate cubic site (**right**).

Department of Civil Engineering and Geological Sciences and Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, IN 46556, USA. E-mail: talbrec1@nd.edu

Published by AAAS

make the experimental measurements difficult. First, precautions are needed to work with radioactive material. Second, the chemical shifts occur over a much wider range than for light elements. Most chemists are accustomed to the narrow chemical shift window of protons that are typically found to yield signals at frequencies shifted by as much as 14 parts per million (ppm) relative to that of a reference standard frequency. Heavy nuclei, particularly those with unpaired electrons, can yield chemical shifts that span thousands of parts per million. Finding a narrow signal that may only have a width of a few parts per million within such a large range is daunting.

Third, the spin of the unpaired electrons couples with the nuclear spin, and the rate at which the nucleus relaxes can be so rapid that it cannot be observed without highly specialized electronics (*2*). It is critical to establish the gyromagnetic ratio (γ) of ^{239}Pu because without this value, the difference in energy between the +½ and −½ spin states cannot be calculated. The resonance frequency of a given nucleus experiencing a specific chemical environment is proportional to both the external magnetic field applied to the sample and its γ value.

There is no simple solution to the spectral width problem, and Yasuoka *et al*. had to sweep a wide range of external magnetic fields to find the signal. The fast relaxation problem was solved by cooling the sample to near liquid helium temperatures of 4 K. Cooling the sample of PuO_2 had two effects. First, it slowed the spin relaxation time to a reasonable value of ~100 s. Second, it ensured a nonmagnetic, singlet ground state for Pu(IV), and the paramagnetic effects were essentially eliminated. It is surprising that the measurements were successful on PuO_2, given that previous studies on the applications of ^{235}U NMR to UO_2 suggested that a sample needed to be antiferromagnetically ordered at low temperatures (*2*) in order to have observable signals. This requirement is apparently unnecessary with Pu(IV) because of its nonmagnetic ground state.

In their supporting information, Yasuoka *et al*. report additional data on a nonstoichiometric sample of PuO_{2-x} (a slightly reduced form of PuO_2). Two distinct chemical signals were observed for ^{239}Pu, which means that it is possible to distinguish between plutonium samples in different oxygen environments (see the figure). Future studies should be able to establish chemical shifts for plutonium in different chemical environments. A key application would be determining the local chemical environments of plutonium in samples where the results of x-ray diffraction are unclear (e.g., mixtures of compounds, slight nonstoichiometries, or disordered samples) or where samples are amorphous. This latter state is exactly what occurs in vitrified nuclear waste, where the plutonium is present within a glass. Thus, ^{239}Pu could be used to determine how plutonium is bound within glasses.

The study by Yasuoka *et al*. provides fundamental information on ^{239}Pu in plutonium dioxide, and on the value of γ and the nuclear magnetic moment, μ_n, for this nucleus. The practical implications of such studies may include the design of advanced materials for storing plutonium for the long term, as well as a technique for probing complex samples that are difficult to deconvolute by other means.

References

1. H. Yasuoka *et al*., *Science* **336**, 901 (2012).
2. K. Ikushima *et al*., *J. Phys. Soc. Jpn.* **67**, 65 (1998).

10.1126/science.1222927

PHYSICS

Room for Just One Photon

Ensembles of cold atoms excited up to Rydberg states can be used to manipulate and control interactions between single photons.

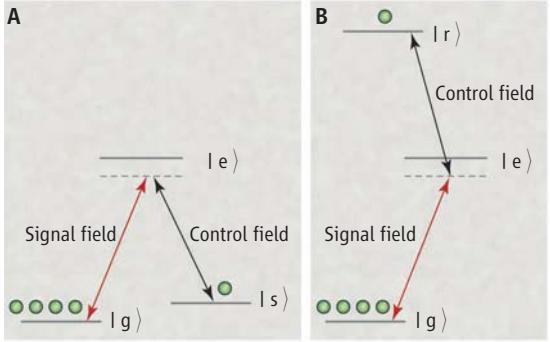
Philippe Grangier

Designing strong, fast, controllable, and switchable interactions between individual quantum systems is a major goal in quantum information processing. This goal, however, presents a difficult challenge because most simple quantum objects, such as photons, generally do not interact with each other. On page 887 of this issue, Dudin and Kuzmich (*1*) demonstrate an interaction with exotic excitations of atomic gases, so-called Rydberg dark-state polaritons (*2*, *3*), that can be turned on and off on a fast time scale and that is strong enough that the presence of just a single photon in the ensemble can prevent other photons to come nearby. Their technique provides a platform to manipulate and control the interactions between single photons.

Institut d'Optique, 91127 Palaiseau, France. E-mail: philippe.grangier@institutoptique.fr

Rydberg atoms, named after the Swedish physicist Johannes Robert Rydberg, are excited atoms with very high values of the principal quantum number n. These atoms are very big, up to micrometer size for n = 100, inflated by a factor of 10,000 times their usual size due to the excited electrons moving far from the nucleus. Such atoms have large electric dipoles (*4*), from which they get the first desirable property: They interact very strongly, at distances of several micrometers, through dipole-dipole interactions with other Rydberg atoms. Charged particles may have even stronger interactions, but Rydberg atoms (which are charge neutral) have another essential quality: Contrary to the charge, the dipole can be turned on and off on submicrosecond time scales by using laser pulses (*3*).

A consequence of these very strong interactions is that when one tries to excite a collection of nearby atoms up to Ryd-



Controlled excitations. In a quantum memory (**A**), a signal field can be stored as a delocalized superposition of two ground states |g⟩ and |s⟩, within an ensemble of atoms. This excitation is called a dark-state polariton ("dark" because the excited state |e⟩ is not involved), and several such polaritons don't interact between themselves. On the other hand (**B**), if the state |s⟩ is turned into a highly excited Rydberg state |r⟩, the corresponding Rydberg polaritons exhibit strong interactions and can mediate interactions between the photons, the result being a nonlinear quantum memory where only one photon can fit.

Published by AAAS

berg states, one atom only will be excited, because as soon as this happens, the interaction will prevent the other atoms from being excited, typically within a range of a few micrometers. This spectacular effect is called Rydberg blockade (*3*), and it has the property of creating quantum entanglement: In a small sample of atoms, only one atom will be excited, but which atom exactly is not known, and this "not known" has to be taken in a quantum sense. Therefore, there will be one excitation only, delocalized over the entire sample. When applied to two neighboring atoms—trapped, for instance, in optical tweezers—this blockade mechanism can create entanglement on demand between atoms. This quantum control can be used to realize quantum gates between atomic qubits (*5*), as demonstrated recently in several experiments (*6*–*9*).

Is there a way to turn this giant interaction between Rydberg atoms into an interaction between photons? This is where dark-state polaritons come into play. Polaritons are mixed light-matter entities (*2*) and have been introduced in the context of quantum memories, as a way to store a photon in an ensemble of atoms, typically as a delocalized quantum superposition of two atomic ground states (see the figure, panel A). It is also possible to make a polariton involving one ground state and one Rydberg state, rather than two ground states (see the figure, panel B). Such a Rydberg polariton will be very short-lived, but that is irrelevant: If a photon, temporarily embodied in a Rydberg state, can interact with other photons brought together in similar states, and then be recreated as a photon, then the goal of effectively inducing controlled interactions between photons will have been attained.

An initial idea was to make a very small sample and to use Rydberg blockade so that only one polariton can be excited. This approach, however, turns out to be not so easy because it is difficult to have an atomic sample small enough to exhibit full blockade, yet big enough to absorb and re-emit a photon with high efficiency. Dudin and Kuzmich developed a different trick based on the polariton being a coherent object: Due to phase matching, the absorbed photon will be re-emitted in the desired direction only if the phase coherence between all the atoms involved in the polariton is preserved during the storage. Then, even if the atomic sample is large enough to accommodate several excitations, the Rydberg interactions will "kill" all polaritons with more than one Rydberg atom, and only the states with one (or zero) excitation will survive (*10*, *11*).

Dudin and Kuzmich measured the single-photon character of the light pulse regenerated from the stored polariton. For low n, the interactions were weak, several polaritons can live in the sample, and the recreated light maintains its classical character; but for high n, only one polariton can fit in the cloud of atoms, and the recreated light turns into a single-photon state (*12*).

The experiment of Dudin and Kuzmich shows that strong Rydberg atom interactions can be used for fast preparation of a single-quantum excitation within a cold atomic gas and that this excitation can be read out as a highly nonclassical light pulse. The results pave the way for building up quantum interactions between photons to achieve photon-photon quantum gates. The technique also allows the study of interacting polaritons to create several entangled states or collective many-body excitations in the context of quantum simulators (*3*). More than one century after their discovery, the strange giant atoms discovered by Johannes Rydberg are still bringing a wealth of surprises.

References

1. Y. O. Dudin, A. Kuzmich, *Science* **336**, 887. 10.1126/science.1217901 (2012).
2. M. Fleischhauer *et al.*, *Rev. Mod. Phys.* **77**, 633 (2005).
3. M. Saffman *et al.*, *Rev. Mod. Phys.* **82**, 2313 (2010).
4. J. M. Raimond *et al.*, *Rev. Mod. Phys.* **73**, 565 (2001).
5. D. Jaksch *et al.*, *Phys. Rev. Lett.* **85**, 2208 (2000).
6. E. Urban *et al.*, *Nat. Phys.* **5**, 110 (2009).
7. A. Gaëtan *et al.*, *Nat. Phys.* **5**, 115 (2009).
8. T. Wilk *et al.*, *Phys. Rev. Lett.* **104**, 010502 (2010).
9. L. Isenhower *et al.*, *Phys. Rev. Lett.* **104**, 010503 (2010).
10. F. Barian *et al.*, *Phys. Rev. Lett.* **108**, 030501 (2012).
11. J. Stanojevic *et al.*, arXiv:1203.6764 [quant-ph] (2012).
12. P. Grangier *et al.*, *Europhys. Lett.* **1**, 173 (1986).

10.1126/science.1222815

CELL BIOLOGY

Ancient Sensor for Ancient Drug

Reuben J. Shaw[1] and Lewis C. Cantley[2]

A common drug has an unexpected effect on a metabolic enzyme that stimulates fat utilization.

Salicylate, a compound found in willow tree bark, is the active breakdown product of aspirin, and has been used for medicinal purposes since ancient times. However, the molecular mechanism underlying the beneficial effects of salicylate has been unclear. On page 918 of this issue, Hawley *et al.* (*1*) show that salicylate activates the cellular metabolic regulator adenosine monophosphate (AMP)–activated protein kinase (AMPK), which stimulates fat utilization in mice.

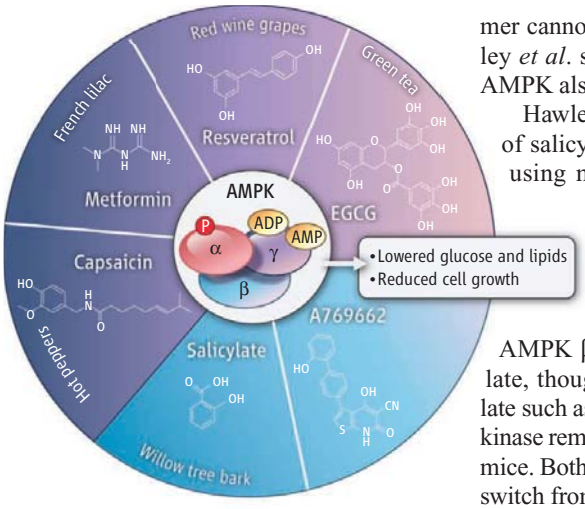
AMPK monitors changes in intracellular energy and is conserved across all eukaryotes. It is activated by a decrease in intracellular adenosine 5′-triphosphate (ATP) concentration, and a concomitant increase in the amounts of AMP and adenosine 5′-diphosphate (ADP). In response to various cellular stresses, AMPK acts as a homeostatic mechanism to restore intracellular ATP by inducing catabolic pathways and inhibiting energy-consuming biosynthetic pathways (*2*). Recent studies have characterized how AMPK is activated by nucleotide concentrations (*3*). Hawley *et al*. demonstrate that salicylate engages an entirely different mechanism to activate the enzyme.

AMPK exists in cells as a heterotrimer composed of a catalytic serine-threonine kinase subunit (α) and two regulatory subunits (β and γ) (see the figure). A number of cellular stresses, such as glucose or oxygen deprivation, cause a decrease in cytosolic ATP and an increase in ADP and AMP, resulting in AMPK activation. The γ subunit contains nucleotide-sensing domains that bind to AMP and ADP (*4*). In a multistep activation process, AMP (or ADP) binds to the γ subunit and induces a conformational change that enhances phosphorylation (and hence activation) of the catalytic subunit. Consistent with this model, AMPK activity in cultured mammalian cells does not respond to insults that lower the concentration of intracellular ATP when the endogenous AMPK γ subunit is replaced with a mutant form that does not bind nucleotide (*5*).

In addition to being activated by nutrient deprivation, AMPK in mammals is also switched on by various xenobiotic and pharmacological agents, many of which act on the mitochondria to lower cellular energy levels. These include antidiabetes agents such as thiazolidinediones (including rosiglitazone) and

[1]Howard Hughes Medical Institute, The Salk Institute for Biological Studies, La Jolla, CA 92037, USA. [2]Beth Israel Deaconess Medical Center, Boston, MA 02115, USA. E-mail: shaw@salk.edu; lewis_cantley@hms.harvard.edu

Published by AAAS



Targeting a metabolic sensor. Many compounds that activate AMPK were first identified as active components of herbal medicines. Several (purple) probably act on mitochondria to increase concentrations of AMP and ADP, which bind the γ subunit. The small molecule A769662 and salicylate (blue), the active derivative of aspirin, activate AMPK through a mechanism involving the AMPK β subunit.

the widely prescribed biguanide compound metformin, all of which are proposed to lower intracellular ATP amounts by inhibiting complex I of the mitochondrial electron transport chain (*2*). Metformin is derived from a compound found first in a medicinal plant, the lilac *Galega officinalis*. By the same proposed manner, other plant products, including resveratrol from grapes and red wine, epigallocatechin gallate (EGCG) from green tea, and capsaicin from peppers, activate AMPK (*2*). In an interesting parallel with those natural products derived from herbal medicines, Hawley *et al*. show that AMPK is also activated by the plant product salicylate.

The connection between diabetes drugs and AMPK activation, and AMPK's control of hepatic and muscle glucose and lipid homeostasis through phosphorylation of metabolic enzymes and transcriptional regulators (*6*), has prompted interest in designing AMPK activators (*7*). However, identifying direct kinase activators is far more difficult than identifying ATP-competitive inhibitors of kinases. To date, A769662 is the only well-established direct small-molecule activator of AMPK (*8*). It was discovered through a screen of over 700,000 compounds for agents that activate purified AMPK in vitro. A769662 does not act by mimicking AMP or ADP binding, but by a mechanism involving the β1 subunit of AMPK (*9*–*11*). The related β2 subunit can restore AMPK responsiveness to energy stress in β1-deficient cells, yet this heterotrimer cannot be activated by A769662. Hawley *et al*. show that salicylate activation of AMPK also requires the β1 subunit.

Hawley *et al*. deciphered which effects of salicylate require AMPK activation by using mice that were genetically engineered to lack the β1 subunit. Although AMPK β2 can compensate for all the critical functions of AMPK β1 in mouse development and physiology, AMPK β2 cannot be activated by salicylate, though all the other targets of salicylate such as IκB kinase and c-Jun N-terminal kinase remain normal and responsive in these mice. Both salicylate and A769662 induced a switch from carbohydrate to fat utilization in mice; this effect was abolished in the AMPK β1-deficient mice. The authors further demonstrate that fatty acid oxidation by salicylate requires AMPK activation, but that other effects on blood glucose and insulin concentrations after prolonged salicylate treatment do not require AMPK.

Future studies should carefully delineate, in different physiological and pathological settings, which effects of therapeutic doses of aspirin and related compounds rely on AMPK activation as compared to other pathways affected by salicylate that are independent of the kinase. This is especially relevant given that aspirin and other AMPK activators such as metformin have been linked to lowered cancer risk (*12*, *13*), and the ongoing examination of whether salicylate derivatives show clinical potential as antidiabetics (*14*).

A full mechanistic understanding of how salicylate activates AMPK through the β1 subunit awaits the isolation of the three-dimensional x-ray crystal structure of salicylate bound to the AMPK heterotrimer, as is also needed for the A769662 compound. That information should make it possible to design activators of this kinase that have higher affinity, greater selectivity, and improved pharmacological properties. Most of the current AMPK activators target the nucleotide-binding pockets of the γ subunit. An interesting possibility is that agents that activate AMPK through its β subunits may synergize with γ-targeted agents. No doubt, this ancient sensor has more mysteries yet to reveal.

References

1. S. A. Hawley *et al*., *Science* **336**, 918 (2012); 10.1126/science.1215327.
2. D. G. Hardie, F. A. Ross, S. A. Hawley, *Nat. Rev. Mol. Cell Biol.* **13**, 251 (2012).
3. D. Carling, F. V. Mayer, M. J. Sanders, S. J. Gamblin, *Nat. Chem. Biol.* **7**, 512 (2011).
4. B. Xiao *et al*., *Nature* **472**, 230 (2011).
5. S. A. Hawley *et al*., *Cell Metab.* **11**, 554 (2010).
6. M. M. Mihaylova, R. J. Shaw, *Nat. Cell Biol.* **13**, 1016 (2011).
7. B. B. Zhang, G. Zhou, C. Li, *Cell Metab.* **9**, 407 (2009).
8. B. Cool *et al*., *Cell Metab.* **3**, 403 (2006).
9. O. Göransson *et al*., *J. Biol. Chem.* **282**, 32549 (2007).
10. M. J. Sanders *et al*., *J. Biol. Chem.* **282**, 32539 (2007).
11. J. W. Scott *et al*., *Chem. Biol.* **15**, 1220 (2008).
12. P. M. Rothwell *et al*., *Lancet* **377**, 31 (2011).
13. G. Taubes, *Science* **335**, 29 (2012).
14. A. B. Goldfine *et al*., *Ann. Intern. Med.* **152**, 346 (2010).

10.1126/science.1223140

CHEMISTRY

Pinning Down the Water Hexamer

Richard J. Saykally[1] and David J. Wales[2]

An experimental study resolves a long-standing controversy about isomeric forms of the water hexamer.

Water comes in numerous forms, from isolated clusters to at least one liquid form and 17 known forms of ice. The detailed characterization of water clusters is essential for developing more accurate and detailed models for describing all these different forms. The comparatively simple hydrogen-bond structures in small water clusters help to elucidate the much more complicated cooperative hydrogen bonding in liquid water (*1*), which is difficult to reproduce quantitatively from simplified models. The water hexamer has received particular attention as the smallest representative of multicyclic three-dimensional hydrogen-bonded structures. On page 897 of this issue, Pérez *et al*. (*2*) report a landmark experimental study that resolves some of the controversies regarding the relative energies of water hexamer isomers.

Calculations of water properties and dynamics typically use a potential energy surface (PES), which gives the energy of a mol-

[1]Department of Chemistry, University of California at Berkeley, Berkeley, CA 94720–1460, USA. [2]Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK. E-mail: saykally@cchem.berkeley.edu; dw34@cam.ac.uk

CREDIT: B. STRAUCH/*SCIENCE*

Published by AAAS

ecule as a function of its geometry. The PES can either be calculated from first principles (ab initio) or with empirically derived functions. Calculations of important water cluster properties, such as tunneling splittings and hydrogen-bond vibrational frequencies, require a PES that can accurately reproduce large amplitude motions far away from local minima. This is a severe test for even the most extensive ab initio calculations. Popular empirical potentials for water can give qualitatively incorrect results even for the cluster structures. For example, the simple point charge extended model (SPC-E) incorrectly predicts the lowest-energy water pentamer and hexamer to be tetrahedral (*3*), and a three-site transferable intermolecular potential (TIP3P) erroneously gives all-atom planar global minima for the water trimer and tetramer.

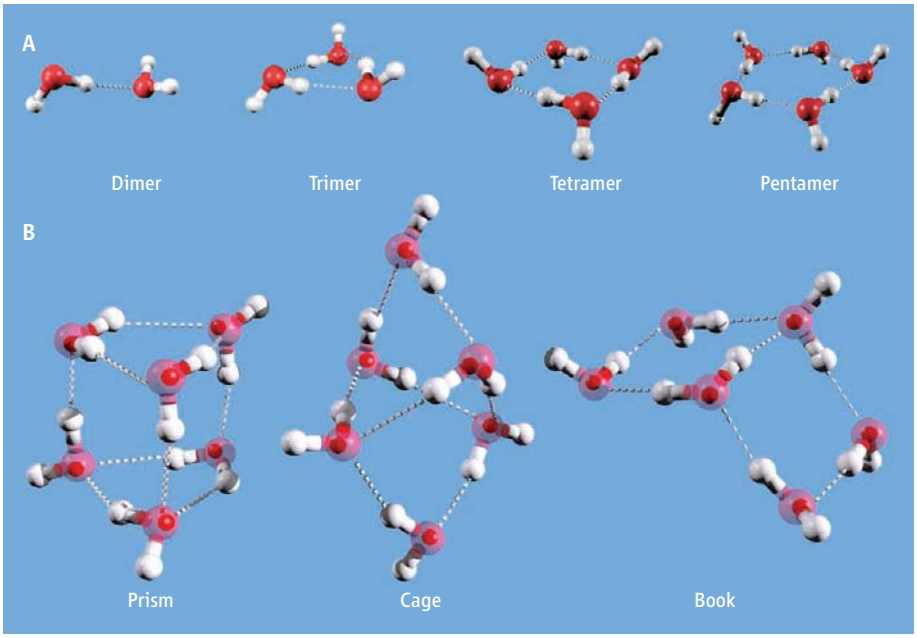
Precise measurement of the hydrogen-bond vibrational frequencies and the associated tunneling has been accomplished for the water dimer through hexamer by far-infrared (terahertz) vibration-rotation-tunneling (VRT) spectroscopy and interpreted from the calculated rearrangement pathways (*4*, *5*). These small clusters sample the canonical hydrogen-bond interactions (electrostatics, induction, dispersion, and exchange repulsion) over a wide range of structures, including the linear dimer, the cyclic trimer, tetramer, and pentamer, and the cage geometry of the hexamer. Experiments and ab initio calculations agree very well through the pentamer. However, there has been controversy regarding the energetic ordering of the low-lying isomers of the hexamer (*2*). It is important to resolve this ambiguity, because the hexamer represents the transition from monocyclic "two-dimensional" to multicyclic "three-dimensional" structures, and global optimization studies indicate that clusters larger than the hexamer all favor such concatenated ring geometries, usually with a number of competing low-energy minima.

According to the tunneling splittings, rotational constants, and torsional vibrations from terahertz VRT spectroscopy, the lowest-energy form of the water hexamer is a "cage" isomer, whereas ab initio calculations and empirical potentials have identified three or four low-energy forms, with most studies identifying a "prism" as the global minimum on the potential surface (*6*). In fact, the highest-level ab initio calculations predict four relatively low-lying cage isomers, but none of them appears to have any low-barrier, single-step rearrangements between permutational isomers (*7*, *8*), which is the usual condition for tunneling splittings to be observed. However, quantum dynamics calculations incorporating only the most important degrees of freedom predict that the lowest vibrational wave functions can be more or less delocalized over the four cage isomers, depending on the intermolecular potential (*7*, *8*). Assignment of the VRT spectrum to a single isomer for the ground vibrational state is thus not inconsistent with the observed tunneling, if the excited states involve different cage isomers. Clearly this system represents a delicate case for theory, in terms of both energetics and dynamics, and both enthalpic and entropic factors must be carefully considered, even at the low temperatures achieved in supersonic expansions.

Pérez *et al*. have now measured the microwave spectrum of the water hexamer. They unambiguously establish the structures and energy ordering of the lowest three isomers (see the figure) by isotopic substitution and temperature variation. The authors also observe rotational spectra of the heptamer and nonamer, but not of the octamer; this is consistent with predictions of a cuboidal octamer structure, which is nonpolar and therefore not detectable in these experiments.

These results were obtained with the powerful broadband Fourier transform method pioneered by Pate and co-workers (*9*) and comprise the first microwave detection of a water cluster larger than dimer, for which spectra were first reported in 1980 (*10*).

The impressive signal-to-noise ratios obtained by Pérez *et al*. promise even more exciting studies of other benchmark water clusters. Along with other new developments, such as the accurate measurement of cluster dissociation energies (*11*), this work will facilitate the long-sought-after "universal" model for water, from which a calculation could accurately predict properties of water in all of its forms (*12*). If such a representation could be found, it would provide a straightforward route to resolving some of the many controversies (*13*) that continue to arise regarding the nature of water.



CREDIT: MATTHEW MUCKLE/UNIVERSITY OF VIRGINIA

Hydrogen bonding in small water clusters. Theory and experiment generally agree regarding the global minimum structures for the water dimer through pentamer (**A**), but not for the hexamer, which represents the transition from monocyclic single donor–single acceptor geometries to multicyclic three-dimensional structures. Pérez *et al*. report the experimental determination of the structures and relative energies of the lowest three hexamer isomers (**B**). Precise characterization of these archetypal examples of aqueous hydrogen bonding advances the quest for an accurate "universal" model of water that applies under all conditions of interest.

References

1. S. Xantheas, *Chem. Phys.* **258**, 225 (2000).
2. C. Pérez *et al*., *Science* **336**, 897 (2012).
3. N. Goldman, R. J. Saykally, *J. Chem. Phys.* **120**, 4777 (2004).
4. F. N. Keutsch, R. J. Saykally, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10533 (2001).
5. T. R. Walsh, D. J. Wales, *J. Chem. Soc., Faraday Trans.* **92**, 2505 (1996).
6. K. Liu *et al*., *J. Phys. Chem. A* **101**, 8995 (1997).
7. D. J. Wales, in *Theory of Atomic and Molecular Clusters*, J. Jellinek, Ed. (Springer, Heidelberg, 1999), pp. 86–110.
8. D. J. Wales, in *Computational Molecular Spectroscopy*, P. Jensen, P. R. Bunker, Eds. (Wiley, Chichester, 2000), pp. 431–456.
9. G. G. Brown *et al*., *Rev. Sci. Instrum.* **79**, 053103 (2008).
10. J. A. Odutola, T. R. Dyke, *J. Chem. Phys.* **72**, 5062 (1980).
11. B. E. Rocher-Casterline, L. C. Ch'ng, A. K. Mollner, H. Reisler, *J. Chem. Phys.* **134**, 211101 (2011).
12. N. Goldman *et al*., *Philos. Trans. R. Soc. A* **363** (2005).
13. A. Nilsson, A. G. M. Pettersson, *Chem. Phys.* **389**, 1 (2011).

10.1126/science.1222007

RETROSPECTIVE

Robert R. Sokal (1926–2012)

An ecologist and evolutionary biologist brought a quantitative approach to classification through statistics and morphometric analysis.

Douglas J. Futuyma

Robert R. Sokal, Distinguished Professor Emeritus of Ecology and Evolution at Stony Brook University, passed away on 9 April 2012 in Stony Brook, New York. He was 86 years old. He was renowned for his contributions to quantitative analysis in biology, especially in ecology, evolutionary biology, and systematics. As a cofounder of "numerical taxonomy," he devised methods for classifying organisms (later applied much more widely) that foreshadowed the development of phylogenetic analysis. He honed a wide variety of statistical tools, especially for analyzing spatially distributed data, and contributed abundantly to analyzing patterns of human genetic variation. His textbook *Biometry,* coauthored with F. J. Rohlf, profoundly influenced training and data analysis in ecology and evolutionary biology.

Robert Sokal was born on 13 January 1926 in Vienna, Austria. In 1939, his family fled the Nazis, escaping by train to Italy and thence by ship to Shanghai, China, joining many thousands of other European Jews. There, he earned a bachelor's degree in biology from St. John's University in 1947, where he performed his first published research (on the anatomy of the head capsule of a dragonfly species). He also met his future wife, Julie Chenchu Yang. His life during these years is described by Stefan Schomann in *Letzte Zuflucht Schanghai* (*Final Refuge Shanghai),* which has been translated into Chinese.

In 1947, Robert moved to the University of Chicago and married Julie a year later. Under the supervision of entomologist Alfred E. Emerson and population geneticist Sewall Wright, he earned a Ph.D. in zoology in 1952 for a statistical study of morphological variation in the aphid *Pemphigus populitransversus* (a subject he would revisit over the next 40 years). He then went to the University of Kansas where he was first an instructor in entomology and then a professor of statistical biology. In 1968, Robert joined the newly created Department of Ecology and Evolution at the State University of New York at Stony Brook, and strongly influenced its development. His career at Stony Brook saw him in the roles of professor, department chair, and graduate program director. He supervised 25 Ph.D. students at Kansas and Stony Brook. After retiring in 1995, he remained an active Distinguished Professor Emeritus in the Department of Ecology and Evolution, continued his research, and stayed involved in the affairs of the university and the U.S. National Academy of Science. Robert attended departmental colloquia until the last year of his life.

Robert Sokal's scientific publications include 13 books (6 translated into other languages) and 206 articles. These include 34 papers on quantitative and ecological genetics of laboratory populations of the fruit fly *Drosophila melanogaster*, the housefly, and the flour beetle, focusing especially on density- and frequency-dependent selection. Perhaps his greatest impact was in developing numerical taxonomy, a quantitative approach to classification to which he devoted two books and 68 articles. Although Robert coauthored one of the first papers on phylogenetic inference, he concluded that classification of organisms could more objectively be based on their quantitative similarity than on their uncertainly inferred evolutionary relationships. His advocacy of this approach embroiled him in intense, sometimes bitter conflict with both traditional taxonomists and the developing school of phylogenetic systematics, or cladistics. His focus on quantitative biology also was evident in more than 50 papers on statistical and biometric methods, especially concerning analysis of spatial data, which he applied mostly to humans and to the aphids that were the subject of his Ph.D. thesis. He published 49 articles on geographic patterns of human genetic variation and its relation to language, ethnohistory, and even cancer rates, focusing mostly on Europe. Robert's broad command of languages (he spoke six) and deep interest in European history contributed to this effort. He used the few dozen available genetic markers to analyze the genetic affinities of various populations, to infer that early agriculture had spread by human migration rather than cultural diffusion, and to show that zones of sharp genetic differentiation correspond to linguistic boundaries.

Robert Sokal served as the president of the Society for the Study of Evolution, the American Society of Naturalists, the Classification Society, and the International Federation of Classification Societies. He was an associate editor of *Evolution* and editor of *The American Naturalis*t. His honors include both Fulbright and Guggenheim awards and the Charles R. Darwin Award for Lifetime Achievement of the American Association of Physical Anthropologists. He was a Fellow of the American Academy of Arts and Sciences and the American Association for the Advancement of Science, and was elected to the U.S. National Academy of Sciences in 1987.

In a professional setting, Robert carried himself with a traditional European formality, was the epitome of organization, efficiency, and attention to detail, and set high standards of performance and productivity for his assistants, students, and especially himself. He expressed his opinions and judgments forthrightly, and by his manner and example portrayed a professional excellence to which younger colleagues (even if perhaps intimidated) could aspire. Robert Sokal was scrupulously honest and fair, and a model of integrity in the face of heated debate and sometimes personal attacks. He was quietly generous and inspired devotion among many of his associates. During the 42 years that I knew him, I first felt honored to have earned his respect, and then gratitude for his strong professional and personal support. In relaxed moments, he was a genial host and an accomplished, humorous raconteur. He appreciated opera and other arts, read widely, and drew sustenance from his Jewish heritage. He developed many close friendships. Most of all, he shared a loving bond with his family—Julie, his wife of 64 years, his two children, and four grandchildren.

10.1126/science.1224101

PHOTO CREDIT: SOKAL FAMILY

Department of Ecology and Evolution, Life Sciences Laboratories, Stony Brook University, Stony Brook, NY 11794–5245, USA. E-mail: futuyma@life.bio.sunysb.edu

Published by AAAS

CREDIT: THINKSTOCK

INTRODUCTION

HUMAN CONFLICT: WINNING THE PEACE

IN THIS SPECIAL ISSUE ON HUMAN CONFLICT, WE CONSIDER THE DEEP EVOLUTIONary roots of violent confrontation. We trace the trajectory of violence and war throughout history, exploring racism, ethnic conflicts, the rise of terrorism, and the possible future of armed conflicts. We also consider our innate capacity to mediate conflict and our ability to achieve—and live in—peace.

Competition and conflict both among and within species, for food or a place to live or a mate, are implicit in the process of evolution and thus intrinsic to our biology. But like many other animals, we are also social beings, and, like them, we have evolved behaviors to avoid the detrimental effects of excessive intraspecies violence. These include ritual singing or fighting displays, acts of submission or conciliation, and simple spatial avoidance seen in diverse species as birds, ants, and our primate relatives.

The groups we consider ourselves to be part of, our "ingroups," are critical to our understanding of "self" and categorization of "other" or "outgroups"; these groups often form the basis for conflict. Groups can have myriad identities. The first were certainly those of kin, and then ethnicity, but as societies grew, groups identified themselves in more complex ways. Our sense of a "group self" influences the way we see and feel about the world around us: We are more likely to empathize with those of our own group and more likely to dehumanize outgroups. Groups based around religious or sacred values can increase ingroup reliance and cooperation through costly rituals, and they can also inspire nonrational sacrifices (suicide bombing, for example), whereas reasoned social contracts seem less able to inspire such fervent loyalty. But although we are ready to see the world as made up of coalitions, it is also true that our perception of "us" and "them" is not primordial because people have constantly shifted affiliations throughout history.

This special section looks at the past and future of conflict and the kinds of data needed to analyze it. However, our interest in conflict must ultimately be to facilitate peace. There are many examples of neighboring societies that do not war against each other, and they appear to have worked in part by expanding the "us." It seems that encompassing a common identity with other groups and transforming outsiders into insiders provides an effective way to overcome conflict. Although some systems (the European Union) still war against outside entities, others (the Orang Asli societies in Malaysia) do not. We hope that understanding how human societies have overcome the odds and developed peaceful relations will help chart a path to a less violent future.

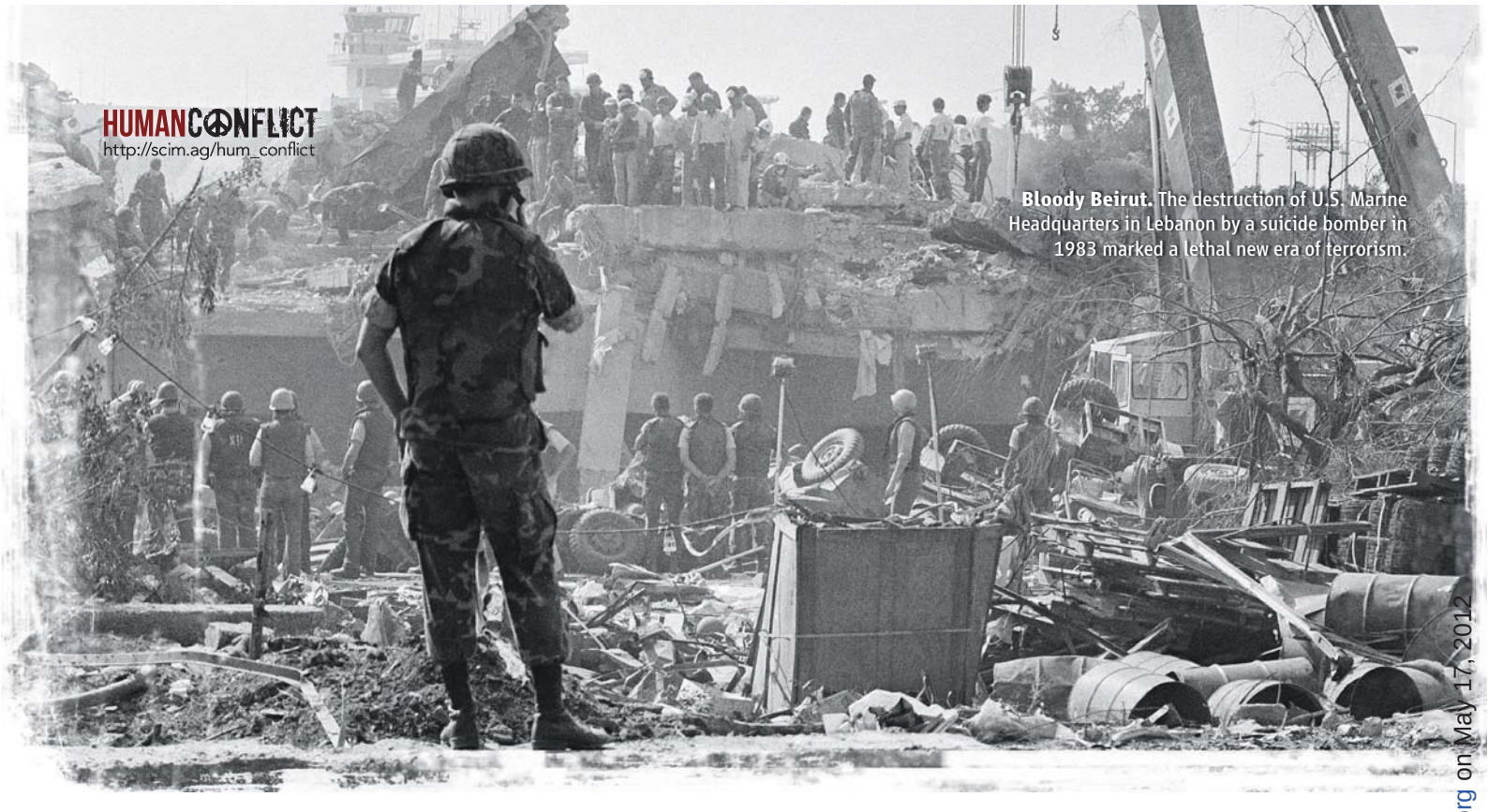
– GUY RIDDIHOUGH, GILBERT CHIN, ELIZABETH CULOTTA, BARBARA JASNY, LESLIE ROBERTS, SACHA VIGNIERI

CONTENTS

News

Reviews

See also Editorial p. 775; Book Reviews pp. 803 and 804; Policy Forum p. 805; and Science *Careers,* Science *Podcast, video, and more at http://scim.ag/hum_conflict.*



HUMANCONFLICT
http://scim.ag/hum_conflict

Bloody Beirut. The destruction of U.S. Marine Headquarters in Lebanon by a suicide bomber in 1983 marked a lethal new era of terrorism.

PARSING TERRORISM

Terrorism research has expanded rapidly since 9/11, shifting its focus from human pathology to the analysis of how rational people interact with violent groups

EARLY REPORTS FROM A CATASTROPHE ON 11 November 1982 in Lebanon's seaside town of Tyre spoke of a suspicious car on the street near a seven-story building. Some people said there was no car. But everyone remembered what happened at about noon: A huge blast ripped through the concrete structure, a base for Israeli forces engaged in a cross-border battle with Palestinians. Scores of Israeli soldiers were killed, along with many other people. It was one of Israel's single greatest military losses.

After an inquiry, Israeli authorities declared it an accident. The explanation: Bottled cooking gas had exploded on the first floor, knocking apart the flimsy walls. But in a clashing version of history, Palestinian militants later claimed that a teenager named Ahmad Qasir died delivering a bomb.

Five months later, a truck loaded with explosives drove into the U.S. Embassy in Lebanon's capital, Beirut, killing 63 people. On 23 October 1983, two more vehicle bombs hit Beirut, one striking a building full of French peacekeepers, killing 58, the other destroying the city's U.S. Marine barracks, killing 241. On 4 November 1983, yet another early morning explosion occurred in Tyre, destroying buildings and killing 29 Israeli soldiers and 32 prisoners. Several more blasts hit other sites in December 1983.

Looking back on the carnage, experts see this period as the dawn of modern, very lethal terrorism. Powered by high explosives and enabled by modern transport systems, much of it has been carried out by young men who knew they would die in the attack. This disturbing form of violence became a global concern after hijackers crashed planes into the Twin Towers in New York City and two other U.S. locations, killing 2996 people on 11 September 2001. Fear surged, and with it a demand for information. People wanted to know who might become a terrorist, what motivates such killers, and what could be done to stop them.

The U.S. government has spent millions of dollars since 9/11 on efforts to answer such questions. It has invited academics to help and encouraged publication in open, peer-reviewed journals. This has produced a bounty of new data sets, case studies, opinion surveys, books, and journal articles. The methods and aims adopted by researchers vary—so much, says economist Eli Berman of the University of California, San Diego (UCSD), who has leapt into this field, that results sometimes seem "like spaghetti." He counts himself part of a new cadre of investigators who strive for quantitative rigor and testable theories.

The study of suicide terrorism in particular has been a "growth industry," says political scientist Martha Crenshaw of Stanford University in Palo Alto, California, a longtime terrorism researcher. She says the work is uneven but becoming more sophisticated. That opinion is echoed by Gary LaFree, director of the National Consortium for the Study of Terrorism and Responses to Ter-

Talking terror. Abu Bakar Bashir in Jakarta, later sentenced to prison for plotting to kill Indonesia's president.

CREDITS (TOP TO BOTTOM): ZOUKI/AP PHOTO; BEAWIHARTA BEAWIHARTA/REUTERS

rorism (START) at the University of Maryland, College Park. "There's been a huge improvement in what we know about terrorist groups," says LaFree, who is proud that in recent years, "we've cracked into most of the major journals."

The work has challenged some long-accepted notions—for example, that terrorists are pathological, driven by religious fanaticism, or spurred by poverty. It's now clear that many terrorists are well-educated and seemingly rational. Economists in particular have explored a disturbing notion, probing terrorism as a form of "altruism" and comparing it to exclusive churches and social aid groups. The studies of cultlike groups are fascinating, Crenshaw says, but don't answer the deepest questions about what makes terrorists tick: "I am not sure that we're closer to understanding" why some groups are violent and others are not.

Give me data

In the 1990s, "there was no one on the planet—at least in the unclassified world—who could tell you how much terrorism was going on," LaFree says. Nobody was collecting and sharing data systematically, although many groups were working on isolated projects. LaFree, who trained as a criminologist, says that just before the 9/11 attacks, he "stumbled onto" a collection of global records describing 60,000 terrorist events assembled by the Pinkerton Global Intelligence Services—on note cards. They would become the core of the Global Terrorism Database, a free digital resource housed at START containing facts and narratives about more than 98,000 events (start.umd.edu/gtd/) that is available to the public. Regularly updated, it has become a key reference.

START was created in 2005 by the Department of Homeland Security (DHS). As one of 12 DHS centers of research excellence, it manages a network of grantees and trains scholars. The Department of Defense and the National Science Foundation together also back an effort known as MINERVA (*Science*, 30 January 2009, p. 576), funding open academic research on terrorism. Many universities have centers of expertise in terrorism.

Scholars have generated hundreds, perhaps thousands, of terrorism studies in the last decade, LaFree says. Researchers have looked for common patterns in the "trajectories" of violent organizations and their leaders in different regions. They have minutely dissected polling data on attitudes to terrorism, examined the link between economic

TERRORISM'S LONG TRAIL

70 C.E. ZEALOTS AND SICARII Jewish militants in Judea assassinated leaders of the Roman Empire and sympathizers.

1000 C.E. "ASSASSINS" Muslim Shia sect in Persia stabbed opposing political and religious leaders.

1870s – 1910s ANARCHISTS – Rooted in the anticzarist movement in Russia, linked to assassinations in several other nations (right an artist's impression of the assassination of Archduke Ferdinand in 1914).

1910s – Today NATIONALISTS – Algerian, Basque, Bosnian, Cypriot, Irish, Kenyan, Serbian, and other groups attack government and civilian targets.

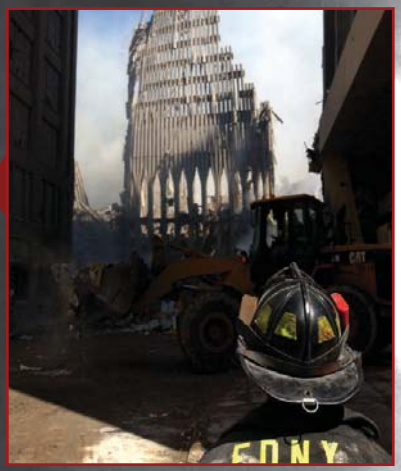
1960s PALESTINE LIBERATION ORGANIZATION – Various attacks aimed at Israel.

1972 MUNICH MASSACRE Eight Palestinians calling themselves Black September broke into the Summer Olympics dormitory, killing 11 Israelis and a local policeman. Five of the eight died onsite; other suspects were tracked down by Israeli agents and killed.

1980s HEZBOLLAH, HAMAS, AND OTHER MUSLIM JIHADIST GROUPS – Bombings and attacks in Israel, Lebanon, Kuwait, and other nations, targeting Israel and supporters.

1987– 2009 LIBERATION TIGERS OF TAMIL EELAM – Set the modern record for suicide attacks (several hundred) in a failed war against the government of Sri Lanka.

1995 TIMOTHY MCVEIGH – U.S. antigovernment loner killed 168 people by bombing a federal building in Oklahoma City.



2001 AL–QAIDA – Muslim jihadists piloted four aircraft in dramatic suicide attacks on U.S. cities.

2008 LASHKAR-E-TAIBA – Pakistani-based terrorists strike in central Mumbai, India, killing 164 people.

2012 MOHAMED MERAH – Muslim jihadist with ties to al-Qaida accused of killing seven people in Toulouse, France.

CREDITS (TOP TO BOTTOM): THE YORCK PROJECT/ DISTRIBUTED BY DIRECTMEDIA PUBLISHING GMBH; AP; KURT STRUMPF/AP PHOTO; U.S. NAVY PHOTO BY PHOTOGRAPHER'S MATE 2ND CLASS JIM WATSON

Published by AAAS

conditions and violence, and analyzed some major government counterterrorist strategies. Some researchers have conducted interviews with terrorists.

Just after 9/11, public attention focused on the links between Islam, radical views, and terrorist killings. Scholars were quick to point out, however, that terrorism has a long and diverse history. Experts such as historian Bruce Hoffman of Georgetown University in Washington, D.C., trace terrorism back 2000 years, defining it as violence or a threat of violence by a nonstate actor with a public or political objective. Jewish fighters known as the Zealots used assassinations as a weapon against the Roman Empire in the 1st century C.E. The anarchist movement of the 19th century killed several European heads of state. In the 20th and 21st centuries, the governments of Britain and Spain, among others, have been battling bomb-wielding nationalists.

It's hard to get evidence on what motivates this violence, but some researchers have tried to do so by asking terrorists. Anthropologist Scott Atran—a prominent investigator at the University of Michigan, Ann Arbor, and the National Center for Scientific Research in Paris—has met with failed, captured, or inactive terrorists in such places as Indonesia, Gaza, and Morocco. The answers they give can seem strangely bland, Atran reports. Many young men told him they joined a violent group because a soccer-playing friend invited them to do so. Recruitment by action-oriented pals is key, Atran says. In testimony to Congress, in scholarly articles (see p. 855), and in a 2010 book, *Talking to the Enemy*, Atran argues that terrorists today are frequently "infected" with the "bug" of radicalism from peers. And that bug is a worldview built on "sacred values" that are not debatable but are linked to violent action. "It is critical to understand how 'sacred values,' such as devotion to God or country or dignity and honor, can motivate" the devoted to act violently, he writes.

Atran doesn't consider fundamentalist religions the source of terrorism; he sees them as repositories of mostly benign values. Although terrorists have sacred values that may be linked to religion, he says, their creeds are action-oriented, alluringly risky, and violent. To counter them, Atran argues, governments need to disrupt the social networks that transmit hostile values and "deradicalize" recruits. Although provocative, such insights rely heavily on interviews and interpretation and are hard to test.

A more down-to-earth idea has been proposed by political scientist Robert Pape of the University of Chicago in Illinois. Zeroing in on suicide terrorism, Pape and colleagues built a database of all the suicide attacks they could find from 1980 to 2001. Looking for commonalities, Pape found that such attacks were growing more frequent and more deadly over time—and that the overwhelming majority (178 out of 188) were linked to one political goal: forcing a democratic power to withdraw from territory it occupied. The evidence, he argued, belied the long-held notion that suicide attackers were mentally unbalanced or inspired by religious dogma. He noted that the most deadly post-1980 terrorists were the Liberation Tigers of Tamil Eelam—a Marxist, atheist group seeking control of Sri Lanka. (The government defeated them in 2009.)

Mapping terrorism. Martha Crenshaw's Web site at Stanford tracks the lineage of violent groups.

Pape and his University of Chicago colleague James Feldman returned to this theme with a larger database that now includes more than 2200 suicide attacks logged between 1980 and 2009. (They've made their data available at cpost.uchicago.edu/search.php.)

Praised by one reviewer as "rigorously empirical," their analysis offers what the authors call "strong confirmation for the hypothesis that military occupation is the main factor driving suicide terrorism." Occupations "account for" 87% of the more than 1800 such attacks examined since 2004, they wrote. They identified a message applicable to any democracy: Don't be an occupier where you're not welcome.

A group of economists found this analysis weak, however, blasting it in a 2008 article in the *American Political Science Review*. Scott Ashworth, Joshua Clinton, Adam Meirowitz, and Kristopher Ramsay of Princeton University argued that Pape had "selected on the dependent variable"—examining only examples of the effect he was trying to study (suicide terrorism) rather than examining all occupations and trying to discern the response. It was a "classic mistake," says Ashworth, who's now at Chicago. Ashworth also sees a contradiction in recent events: There have been suicide attacks in Pakistan, where there is no foreign occupation.

Pape responds that, in a "slight modification" of his theory, he has categorized Pakistan as being "indirectly occupied" by the U.S. military. And this, he says, can also lead to suicide terrorism. Pape rejects the criticism of his methods, saying that in books he published in 2005 and 2010, he did exactly what the critics wanted: "I collected not a sample, but the entire universe" of military occupations and suicide attacks. He predicts that suicide attacks will "fade" in the next 5 years as U.S. troops leave Afghanistan.

Lethal dreams. A bomb ignites in 2009 during a Muslim ceremony in Sri Lanka (*left*), set off by a member of the Liberation Tigers of Tamil Eelam (*right*). This Marxist outfit, now defunct, took credit for more suicide bombings than any other group.

CREDITS (TOP TO BOTTOM): ROD SEARCEY/ROD@RODSEARCEY.COM; REUTERS TV/REUTERS; ANURUDDHA LOKUHAPUARACHCHI/REUTERS

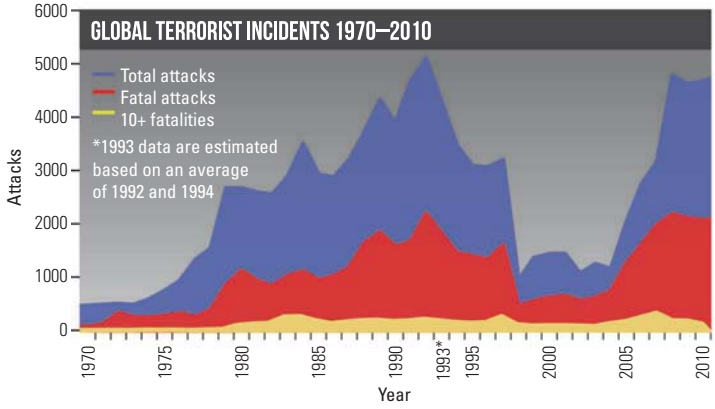
Economy of terror

Economists have been delving into terrorism, too. They're probing questions such as whether the poor are more likely to carry out violent attacks (the consensus is that they are not), the internal dynamics of radical groups, and the effect of terrorist attacks on public opinion. Some of their findings are unsettling. For example, a 2010 paper by economists Esteban Klor and Eric Gould of the Hebrew University of Jerusalem compared polling data from Tel Aviv and Haifa before and after terrorist attacks that took place at different times. The researchers found consistent results: In both locations, more people favored compromise with the Palestinians after attacks than before.

The best studies, says Ethan Bueno de Mesquita, a political scientist at the University of Chicago who recently surveyed the field, analyze "fine-grained" information about specific people and events. He says "a small group of scholars" is doing most of this work, beginning with a well-defined theory and gathering "microlevel" data, such as details of attitude changes before and after an event in a specific location. Focusing in this way makes it easier to search for cause and effect, de Mesquita says, but there's a tradeoff: Conclusions tend to be fine-grained, too.

Berman, research director at UCSD's Institute on Global Conflict and Cooperation, is an economist who has also tried to develop a general hypothesis and gather micro-level data. His starting point is a bit "jarring," he concedes: Terrorists see themselves as altruists and are "fairly rational actors." The assumption is based on research by Israeli psychologist Ariel Merari and others, who have "gone into cells" to interview suicide attackers who didn't succeed or were caught, Berman says. He used interview data and records of attacks in Israel and Israeli-occupied territories to zero in on the deadliest groups. He argues that they rely on the same loyalty-boosting methods used by religious sects and mutual aid organizations.

Berman's analysis is an outgrowth of work by economist Laurence Iannaccone on the dynamics of religious sects. A salient aspect of all sectlike mutual aid groups, according to Iannaccone and Berman, is that they raise barriers against newcomers and require big sacrifices from members. The advantage for aid groups is that they can exclude free riders; terrorist groups gain by weeding out potential defectors. The most restrictive sects—those that demand hours of devotion, impose dress codes, and rigidly control personal interactions—bind members in a kind of family that makes the group more "effective," according to Berman. In the economy of terror, the more effective groups kill more people.



Unsteady rise. A change of methods partly explains the dip in records at the Global Terrorism Database at the University of Maryland.

Berman tested this idea several ways: for example, by analyzing all suicide attacks in Israel, Israeli-occupied territories, and Lebanon between 1981 and 2003, as he describes in his 2009 book *Radical, Religious, and Violent*. Using the number of fatalities per attack as a "lethality" index, Berman found that faith-based providers of social services were the most dangerous. Two out of seven groups stood out: Hamas and Hezbollah, topping the list with 63 and 44 fatalities per attack. The Palestinian Islamic Jihad ranked third, at 37 fatalities per attack, followed by others ranging from 31 to 1. Berman says data from Iraq follow the same pattern, linking high lethality and religious radicalism.

Dead certainties. Anthropologist Scott Atran studies "sacred values" held by terrorists.

de Mesquita says Berman's studies are "the most interesting work about terrorism and religion," adding that they're not really about religion, but about methods to boost group commitment and avoid infiltration. Many questions remain unanswered, however. For example, why are some sectlike groups benign while others are terribly violent? Berman concedes that he doesn't have an answer.

There are many gaps in terrorism research, says social psychologist and expert on radicalization, Clark McCauley of Bryn Mawr College in Pennsylvania. He is concerned, for example, that by rejecting the idea that pathology lies at the heart of terrorism, "maybe we went too far" in looking for explanations in normal group behavior. McCauley recently studied 200 criminal cases linked to al-Qaida and discovered that 40 involved "lone wolf" terrorists. A recent example is Mohamed Merah, who was killed in a shootout with police in March after gunning down citizens and paratroopers near Toulouse, France. Another suspected of acting alone is U.S. Army Major Nidal Malik Hasan, accused of an attack at Fort Hood, Texas, in November 2009 in which nine people were killed and 30 injured. This suggests that lone actors are a significant subcategory deserving a second look, he says. After preliminary research, he has a hunch that these people may not be distinguished by an emotional deficit but by a surplus of empathetic feelings for perceived victims—and a willingness to act on their behalf.

McCauley is keenly aware that there are few certainties in the study of terrorism—so much so that he considers himself "a connoisseur of holes." The biggest hole he sees is the failure to keep good data on governmental attempts to counter terrorism and their consequences. "We don't know when we instituted a new initiative and we don't know when we changed it, or stopped doing it," he says. He says the demand for objective data is growing.

Like McCauley and other terrorism researchers, Crenshaw acknowledges that the gaps can be frustrating. Even some widely reported events—such as the 1982 blast in Tyre—remain murky, their cause disputed. But Crenshaw says the field now has excellent public databases and a growing community of investigators. We know a lot more about terrorism than we did a decade ago.

–ELIOT MARSHALL

CREDITS (TOP TO BOTTOM): (SOURCE) GLOBAL TERRORISM DATABASE; RICH DAVIS

IN BATTLE

TRIBAL ROOTS IN SOUTH SUDAN

MURDER WAS ON THE AGENDA FOR NEW YEAR'S DAY IN Pibor, a remote town in Jonglei State in South Sudan. Isolated by bad roads and poor communication, Pibor is home to a people known as the Murle, feared by their neighbors and also often attacked by them. In late 2011 and early 2012, a spokesperson for a group of the neighboring Lou Nuer tribe made the ultimate threat. *The New York Times* and other papers quoted him as saying that they intended to "wipe out the Murle tribe on the face of the earth."

A vigilante army of more than 6000 young Nuer men mobilized in late December, armed with automatic rifles and knives, calling themselves the White Army. They accused the Murle of constantly stealing cattle and kidnapping women and children—a charge frequently made across Jonglei State, homeland of the Murle, Nuer, Dinka, and other tribes. The incitement to attack, the White Army said, was a Murle raid against the Nuer in August 2011 in which 600 people were killed, women and children were kidnapped, and tens of thousands of cattle were stolen.

At the end of December, the vice president of South Sudan's national government, Riek Machar, met the war-bound Nuer to plead for peace. He failed. The government stationed 400 troops near Pibor but did not intervene. United Nations peacekeepers, with a force of 400 in the area, also kept their distance. As the White Army marched on Pibor, burning huts along the way, U.N. officials advised people to flee.

The town emptied, and more than 50,000 people hid in the bush without food, fresh water, or shelter for a week. When the Nuer began to withdraw on 3 January, the Murle side said that 3000 people had died or been killed and hundreds of thousands of cows taken. A U.N. official said the number of deaths was "nowhere near" that high, although still unknown.

Like January's death toll, the start of this violence is hard to pin down. Ask Gai Bol Thong, a Nuer partisan in Seattle, Washington, and an advocate of the White Army: He says the attacks and revenge strikes between the Murle, the Nuer, and others in Jonglei "began before the liberation of Sudan" from colonial rule in 1956. Such raids may have been going on for centuries.

Conflict has long been part of the all-embracing cattle economy of Jonglei. The animals are food (there's little agriculture), the embodiment of wealth, and a prerequisite for fatherhood. To start a family, a man must have cattle to buy a wife. The cattle are herded by armies of young men, who clash with others as they move cows to fresh feeding areas, often through neighbors' lands. Cattle herders may also be cattle takers, and rustling is a time-honored way of securing a dowry. Children are part of this economy, too: Boys may be abducted to work as herders by a competing tribe, and girls may be taken to be exchanged for cattle. The Murle are not the only ones engaged in this trade.

These patterns may be centuries old, but the impact of fighting has intensified since the 1990s, according to many Sudan analysts. Factions in a long civil war between northern and southern Sudan distributed automatic weapons far and wide; these are now in the hands of uncontrolled militia. The government of South Sudan, just 10 months old, hasn't been able to confront the anarchy at home, in part because it is still fighting on the border with Sudan. (Ethnic battles appear at least as severe within Sudan, which is governed by the world's only sitting head of state charged with genocide by the International Criminal Court, for attempting to eradicate nomadic tribes in Darfur.)

For now, South Sudan's leaders have chosen a simple remedy to combat violence. In April, they launched a campaign to confiscate weapons from civilians. Hundreds have been collected already, but experts say this is not a viable long-term strategy. John Young, a Canadian political scientist who tracks the conflict in Jonglei for the Carter Center in Atlanta, notes that it has been tried before. People can now get guns when they want them. Disarmament, he wrote in 2010, is "unlikely to prove enduring" because it does not address the underlying causes of conflict "nor the deep poverty of the people which has made abductions a lucrative business."

The solution, as many investigators have concluded, is to provide real public security, improve roads and communication, stimulate the economy, and ensure fair and honest governance.

–ELIOT MARSHALL

CREDITS (TOP AND BACKGROUND): ISAAC BILLY/U.N. PHOTO

Fear of "them." Members of outgroups can spark automatic, implicit prejudice.

ROOTS OF RACISM

Humans everywhere divide the world into "us" and "them." Why are we so tribal?

YOU'RE ALONE IN A DARK ALLEY LATE AT NIGHT. Suddenly a man emerges from a doorway. If you are a typical white American and he is a young black man, within a few tenths of a second you will feel a frisson of fear as your brain automatically categorizes him. Your heart beats faster and your body tenses.

In this event, nothing happens. He glances at you and moves away. You walk on, feeling foolish for fears based merely on his membership in a racial group.

Online

sciencemag.org

Podcast interview with author Elizabeth Culotta (http://scim.ag/6083_hc1).

Tension and suspicion between groups—whether based on racial, ethnic, religious, or some other difference—fuel much of the world's violence. From the enduring feuds of the Middle East and Northern Ireland, to the vicious raids of South Sudan, to the gang warfare that plagues American cities, even to bullying in schools and skirmishes between fans of rival sports teams, much of the conflict we see today erupts because "we" are pitted against "them."

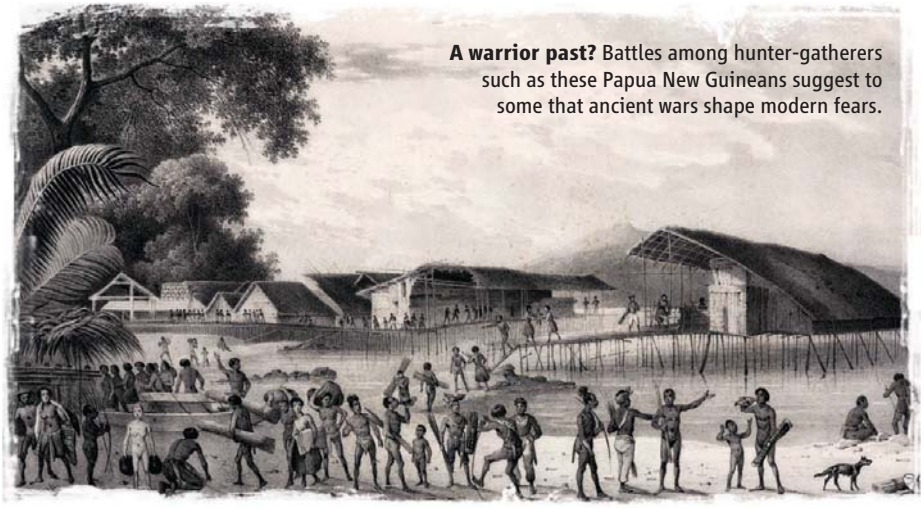
Some of the prejudice behind these conflicts is not conscious: Your fear spiked in that dark alley before your conscious brain had even registered the young man's skin color. This prejudice apparently stems from deep evolutionary roots and a universal tendency to form coalitions and favor our own side. And yet what makes a "group" is mercurial: In experiments, people easily form coalitions based on meaningless traits such as preferring one painter over another—and then favor others in their "group," giving them more money in games, for example. "In arbitrarily constructed, meaningless groups with no history, people still think that those in their ingroup are smarter, better, more moral, and more just than members of outgroups," says Harvard University psychologist James Sidanius.

A wide and deep literature has explored these innate biases in the 40 or so years since they were first discussed. Now researchers have begun to ask *why* humans are apparently primed to see the world as ingroups and outgroups. What factors in our evolutionary past have shaped our coalitionary present—and what, if anything, can we do about it now?

Several avenues of research are probing the origins of what many psychologists call ingroup love and outgroup hate. Researchers are testing the implicit biases of young children and even primates, and devising experiments to ratchet bias up and down. Evolutionary researchers are trying to parse the group environments of our ancestors and are debating just how big a selective pressure came from outgroup male warriors. "The origin of all this is the all-consuming question of the past few years," says Harvard psychologist Mahzarin Banaji.

Group love

For many researchers, our cruelty to "them" starts with our kindness to "us." Humans are the only animal that cooperates so extensively with nonkin, and researchers say that, like big brains, group life is a quintessential human adaptation. (In fact, many think big brains evolved in part to cope with group living.) Studies of living hunter-gatherers, who may represent the lifestyle of our ancestors, support this idea. Hunter-gatherers "cooperate massively in the flow of every imaginable good and service you can think of," says anthropologist Kim Hill of Arizona State University (ASU), Tempe, who has studied



A warrior past? Battles among hunter-gatherers such as these Papua New Guineans suggest to some that ancient wars shape modern fears.

CREDITS (TOP TO BOTTOM): © JUSTIN MINNS/ALAMY; SSPL VIA GETTY IMAGES

Published by AAAS

Religion, not race. Groups divide along many axes, including religion, as when protesters and police clash in Northern Ireland.

hunter-gatherers for 35 years. "Anything you need in daily life, the person next to you will lend you: water, sticks for firewood, a bow and arrow, a carrying basket—anything."

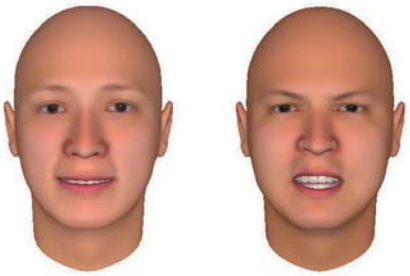
Thus the group buffers the individual against the environment. "Our central adaptation is to group living," says psychologist Marilynn Brewer of the University of New South Wales in Sydney, Australia. "The group is primary."

When the ingroup is loved, by definition there must be a less privileged outgroup. "One can be expected to be treated more nicely by ingroup members than by outgroups," as Brewer put it in a seminal 1999 paper. "It is in a sense universally true that 'we' are more peaceful, trustworthy, friendly, and honest than 'they,' " she wrote.

If groups compete for territory or resources, favoring the ingroup necessarily means beating the outgroup and can escalate into hostility, Brewer notes. Several other researchers (see p. 876) have recently argued the reverse: that over time, hostilities between groups fostered ingroup love, because more cooperative groups won battles. Whichever came first, researchers agree that outgroup hate and ingroup love may have spurred each other.

In the United States and some other countries, the sharpest division between groups is often racial. But researchers agree that it's not that white people have evolved to be suspicious of black people per se or vice versa. Such an evolved prejudice could only arise as the result of frequent negative interactions between races in the past, explains anthropologist Robert Boyd of the University of California, Los Angeles. But thousands of years ago, people didn't cross continents to meet each other. "In the distant past, we had very little experience interacting with people who were physically very different from us," Boyd says. "That's only since 1492. Ethnic distinctions, however, are presumably quite old." Thus racial prejudice is a subset of a much broader phenomenon. "This is not just about racism," says psychologist Susan Fiske of Princeton University.

The targets of outgroup prejudice vary from culture to culture and over time—Sidanius refers to them as "arbitrary set" prejudices. In Sri Lanka, it may be Tamils; in Northern Ireland, Catholics or Protestants; in India, the Untouchables. Fiske notes that the world over, the greatest prejudice is often aimed at people without an address, such as gypsies and the homeless. Whoever the target, we have a psychological system that prepares us to "learn quickly, in whatever cultural context we're in, what are the cues that discriminate between us and them," says psychologist Mark Schaller of the University of British Columbia, Vancouver, in Canada.



Whose group? Taiwanese toddlers thought racially ambiguous faces like these were Taiwanese when smiling and white when looking angry.

This doesn't mean that prejudicial behavior is inevitable, Schaller says. "These prejudices tap into very ancient parts of our minds, and it's happening at a very quick, automatic level," he says. "But we have recently evolved parts of our brains that allow us to engage in slower, more rational thought. When I experience that fear in a dark alley, it may take me another half-second for a more rational thought to kick in, but I'll get there, if I have the motivation and means to do so."

Shoot or don't shoot

Psychologists have become master manipulators of prejudice in the lab, with clever experiments that reveal underlying biases. In the Implicit Associations Test, for example, people are asked to rapidly categorize objects and faces; the pattern of mistakes and speed shows that people more quickly associate negative words such as "hatred" with outgroup faces than ingroup faces. "It takes significantly longer to associate your ingroup with bad things and the outgroup with good things," Sidanius says. In disturbing tests using a video game, people looking at a picture of a person carrying an ambiguous object are more likely to mistake a cell phone for a gun and shoot the carrier if he is an outgroup male.

This type of bias shows up in all cultures studied and in children; it also appears in people who say they are not prejudiced and who work consciously for equality. "This is in every single one of us, including me," Banaji says.

It starts young. In work in review, Yarrow Dunham of Princeton, Banaji, and colleagues found that Taiwanese toddlers assumed that a smiling racially ambiguous face was Taiwanese, but a frowning one was white; white, American 3-year-olds similarly preferred their ingroup.

Even our primate cousins categorize others into ingroups and outgroups. Chimps obviously have outgroup bias: They sometimes band together and attack and kill members of other troops. Last year, a study managed to uncover primates' implicit expectations of "us" and "them" for the first time.

Laurie Santos of Yale University, working with Banaji and others, adapted a psychological test for rhesus macaques, group-living monkeys whose lineage diverged from ours about 25 million to 30 million years ago. The researchers assumed that the macaques would stare longer at outgroup faces, who might be more dangerous, or at groups of photos that paired things they liked with outgroup faces they didn't like. In a series of experiments pub-

CREDITS (TOP TO BOTTOM): STRINGER IRELAND/REUTERS; YARROW DUNHAM

lished in the *Journal of Personality and Social Psychology*, Santos and colleagues found that macaques looked longer at photos of outgroup members. They also looked longer at photos of outgroup members next to pictures of fruit, and at photos of ingroup members with spiders and snakes.

Seeing such apparent bias in primates suggests it is evolutionarily ancient. This "coherence of results across species and ages is satisfying," Banaji says, and tells us that outgroup bias is "core to our species."

While these researchers probe bias in different populations, others are exploring how to manipulate it. Our attitudes toward outgroups are part of a threat-detection system that allows us to rapidly determine friend from foe, says psychologist Steven Neuberg of ASU Tempe. The problem, he says, is that like smoke detectors, the system is designed to give many false alarms rather than miss a true threat. So outgroup faces alarm us even when there is no danger.

Neuberg and Schaller have studied what might turn this detection system up and down. When you feel threatened, you react to danger more quickly and intensely; people startle more easily in the dark. That's why prejudice rears its head in a dark alley rather than a well-lit field. In a variety of studies, Neuberg, Schaller, and their colleagues have manipulated people into feeling unconsciously more fearful or confident and found that measures of outgroup bias respond. Canadians taking tests in the dark rated Iraqis as less trustworthy and more hostile than other Canadians. Sinhalese in Sri Lanka stereotyped Tamils as more hostile after being primed with a geographic context that made them feel outnumbered. And white undergraduates were more likely to misperceive anger on the faces of black men—but not whites—after watching a scary scene from the movie *The Silence of the Lambs*. Schaller adds that some people seem to go through life more cognizant of threats than others, and that prejudice is more easily intensified in these people.

These findings can inform real-life tragedy, researchers say. On the evening of 26 February, a Hispanic man named George Zimmerman shot an unarmed black teenager, Trayvon Martin, in a gated community in Sanford, Florida. Zimmerman told police that he followed Martin suspecting criminal activity, was attacked, and fired in self-defense. Researchers cannot speak to what happened that night. But when Zimmerman first spotted Martin, the situation was a "perfect storm" for triggering feelings of vulnerability and implicit prejudice, Neuberg and Schaller say.

It was dark and raining. Martin, 17, though slender, was tall. Zimmerman, 28, was quite alert to crime in his neighborhood; he had started the neighborhood watch. Martin was young, male, and black, an outgroup stereotyped as dangerous by whites and Hispanics in the United States. "We would predict that under those circumstances this kind of thing would happen more often," Neuberg says.

On the lookout. Macaques stared longer at photos of the faces of outgroup members than at ingroup faces.

CREDIT: LAURIE SANTOS

If certain situations turn implicit prejudice up, can it be turned down? Schaller notes that making people feel safer can moderate this bias, whether through specific priming or more generally with lower crime rates or a better economy. To unconsciously prime her own mind, Banaji has created a screen saver that displays stereotype-smashing images. Other researchers say that deliberately engaging the slower conscious mind may help. For example, in addition to skimming all job applications quickly, a manager might read the files of minority applicants with care.

Men in the crosshairs

Martin is a good example of how outgroup prejudice falls hardest on men, Sidanius says. He argues that this has specific evolutionary roots: Because it is men who typically make war, it was outgroup men who attacked our ancestors; it was also men who were more likely to be killed in combat. "Back in the Pleistocene, outgroup males really were dangerous," he says.

If natural selection has shaped our minds to be wary of outgroup males, then they should face more prejudice than outgroup women, says Sidanius, an African American who himself was the target of hate crimes as a young man. He and colleagues have assembled a devastating catalog showing how this is true for black men in America. As compared with black women, black men are more likely to be victims of hate crimes, receive harsher jail sentences for comparable offenses, pay more money for cars—the list goes on and on. Data suggest that West Indian and South Asian men in the United Kingdom face similarly disproportionate bias, Sidanius says.

Building on these ideas, in March, Melissa McDonald of Michigan State University in East Lansing and colleagues proposed what they called the "warrior male hypothesis," arguing that natural selection has shaped men's minds, more than women's, toward belonging to coalitions. They predict that men are more prejudiced than women, and some data show this.

But others aren't so sure that intergroup war was a prominent feature of our prehistory (see p. 829). Foragers depend on far-flung networks to gain access to the social and natural resources of others, notes anthropologist Polly Wiessner of the University of Utah in Salt Lake City. The !Kung of Africa, whom Wiessner studied for decades, "may travel for hundreds of miles to visit exchange partners in less familiar areas, with no fear of unknown males," she says.

Whether or not humans have evolved to fear outgroup men per se, researchers agree that we are prone to categorize and sometimes fear outgroups. "What we're arguing is a natural preference for drawing ingroup-outgroup boundaries. It can be race, religion, nationality, dialect, or arbitrary set differences," Sidanius says. "But once those boundaries are drawn, people like to discriminate across them."

–ELIZABETH CULOTTA

Published by AAAS

IN BATTLE

PREENING THE TROOPS

"KEK-EK-EK-EK! KEK-EK-EK-EK!" THAT'S A CALL TO ARMS FOR the green wood hoopoe, a 44-centimeter-long South African bird that forms alliances among a breeding pair and up to 10 helpers. As soon as one bird starts calling, others join in, in a display of solidarity that can be heard a kilometer away.

Listen closely and not one but two groups typically sound off, taking turns to see which one calls loudest and longest. The birds depend on these "rallies" to settle disputes, which today represent one of the best documented examples of group conflict management in animals other than primates, says behavioral ecologist Andrew Radford of the University of Bristol in the United Kingdom. "Nonprimates such as birds … offer many good opportunities to generate hypotheses about intergroup behavior," says anthropologist Richard Wrangham of Harvard University.

Just as humans shower hometown athletes with praise at pep rallies and postgame celebrations, wood hoopoes boost in-group cooperative behavior by preening helpers more before and after rallies. This avian behavior shows that it doesn't require a primate brain to reinforce social bonds in the face of an outside threat.

In wooded ravines bordering rivers, the birds set up permanent territories that stretch about 1.5 kilometers long but are typically just several hundred meters wide. Groups of two to 12 birds face off whenever one crosses or nears a territory boundary. Losers retreat, allowing winners to barge in to feed on insects and other invertebrates in the treetops, and to nose around in roosting holes before returning to their original territory.

When groups face off in a display like the wood hoopoe rallies, the number of individuals can matter more than the size or stamina of any one member (see p. 838). Roaring in lions, for example, indicates a pride's size and its fighting ability.

Radford created artificial contests by playing prerecorded rallies to wood hoopoes. During a rally, the birds in a group sit close together and rock back and forth, calling for about 10 seconds per bout. The calls are coordinated, with each bird cackling in turn "like a well-oiled chorus," Radford says. Rival groups take turns sounding off.

Radford found that a group cackled longer if the challenging band was larger. Longer responses even by small groups may be an attempt to disguise their small number, he says. Some contests were short, less than 5 minutes, and usually won by resident birds no matter their group size. But other battles stretched more than 15 minutes, with calling bouts getting longer as residents matched intruders' efforts. Larger groups usually won such extended contests.

Researchers have long suggested that conflict between groups strengthens the cooperation and solidarity within them; some say this has shaped humans' tendency to form coalitions (see p. 825). Experiments have borne out this idea, such as those pitting groups of college students playing a cooperation game against each other. Radford investigated whether the same held true in the green wood hoopoes, looking at friendly gestures among birds after rallies. "Looking at the aftermath, virtually no one had done that," he says. "But it has clear parallels with human battles."

Three cheers. Green wood hoopoes foster group cohesion as they prepare to match calls against other groups.

To measure affiliative behavior, he noted incidents of allopreening in which one bird picks at the feathers of another, often to remove parasites. He watched 12 groups of birds, charting rallies as well as who groomed whom and when. Short contests didn't lead to a change in preening, but long conflicts, particularly when lost, spurred more preening of group members, particularly by the dominant pair, he reported in 2008. He thinks the postconflict preening helps reduce social stress caused by the conflict. The dominants may use grooming not only to soothe the helpers, who do the lion's share of cackling during contests, but also to "enhance the helper's participation in the future," similar to how primates groom one another to strengthen bonds, he says.

"This important finding indicates that wood hoopoes have evolved a partial solution to a collective action problem" of some group members working harder than others, Wrangham says. "It's the kind that is worth looking for in primates."

But Radford also noticed that body preening increases when the birds simply enter a part of their territory where past conflicts occurred. When the birds approach their borders, "there's a preemptive change in behavior," he notes. The birds weren't just more anxious, because self-preening—considered a sign of increased stress—did not increase, he reported online on 7 July 2010 in *Biology Letters*. This finding parallels what is seen in soldiers, who bond more when they enter combat zones than when stationed back from the front lines. "If we are finding these links in birds and not just among humans, it suggests these links could be very important over evolutionary time," Radford says.

–ELIZABETH PENNISI

THE BATTLE OVER VIOLENCE

It takes a village. Violence appears in every culture but becomes more complex as societies develop.

Under the long shadow of Rousseau and Hobbes, scientists debate whether civilization spurred or inhibited warfare—and whether we have the data to know

WITH ITS WORLD WARS, GENOCIDES, AND innumerable revolutions and civil wars, the 20th century was the bloodiest in human history. World War II alone left some 60 million dead—2.5% of the world's population, or the total number of people who lived in Europe during the Middle Ages. Yet a group of researchers argues that complex industrialized societies, even Nazi Germany or Stalin's Russia, are far safer places to live than among smaller groups of hunter-gatherers or farmers, where tribal feuds and homicide typically felled more than 10% of the population.

"This is the paradox," says Azar Gat, a military historian at Israel's Tel Aviv University and author of a major study on the subject. "Mortality was higher before the state appeared." Perhaps the most prominent advocate of this view is Harvard University neuroscientist Steven Pinker, who argued in his 2011 book *The Better Angels of Our Nature* that complex societies have spurred a clear and measurable decline in the rate of violence compared to the bad old days when humans lived primarily in more intimate, small groups. With the centralized state, he writes, "came a more or less fivefold decrease in rates of violent death."

Many archaeologists and anthropologists agree that the odds of dying violently are lower in a modern nation today than in medieval Europe, a comparison prominent in Pinker's book and one that draws on the relatively plentiful records of the past 5 centuries. But a number of researchers dispute the more general assertion that violence decreases as a society becomes more complex. This theory, they maintain, ignores yawning gaps in data as well as the enormous diversity of rates and forms that violence takes from one society to the next.

"The variability across cultures in how violence is used is staggering," says Debra Martin of the University of Nevada, Las Vegas, who specializes in ancient health and violence. Adds Henry Wright, an archaeologist at the University of Michigan, Ann Arbor, who has studied violent death in the ancient Near East, China, and Africa: "I don't have the statistics, and I don't think anyone else does either."

Beneath the argument over numbers is a controversy reaching back several centuries. Swiss-born French philosopher Jean-Jacques Rousseau argued in the 18th century that complex society brought about greater inequality, oppression, and fear. Thomas Hobbes, a 17th century English philosopher, argued that life without the social order and a strict hierarchy was "solitary, poor, nasty, brutish and short." Whether civilization corrupted humans or saved them has made sparks among scholars ever since.

Early and mid-20th century studies of ancient people seemed to confirm a more Rousseauian view in which scattered populations, minimal technology, and ample game limited human violent conflict in the distant past. Cave paintings in Europe from about 40,000 to 10,000 years ago portray hunting of animals but not human-on-human conflict. Archaeologists found little evidence of murder and organized violence before the military empires of the Near East sprang up 4000 years ago. Studies of living hunter-gatherer tribes in the first half of the 20th century appeared to show low rates of violence: American anthropologist Margaret Mead concluded in 1935 that in the Arapesh tribe of New Guinea, "both men and women are naturally maternal, gentle, responsive, and unaggressive." And initial primate research found fewer violent tendencies in humanity's nearest cousins.

This Rousseauian perspective began to lose favor a half-century ago. Early Neolithic cave paintings in Spain recorded in the 1980s show humans shooting arrows at one another. Primatologists discovered that warfare and murder are not unusual among chimpanzees. And more intensive anthropological work began to shed light on a more violent side of small-group society.

In 1996, anthropologist Lawrence Keeley of the University of Illinois, Chicago, published *War Before Civilization: The Myth of the Peaceful Savage*, based on a wide range of data from prehistoric sites, modern hunter-gatherers, and other groups living outside established states. He concluded that more than 90% of human groups engage in war, including small-scale groups. For those people living outside states, Keeley estimated that the average annual rate of death in warfare was 524 per 100,000 people—twice that of the famously warlike Mesoamerican Aztecs in the 16th century. By contrast, even during the bloodiest years of World War II, Russia and Germany had violent death rates of about 140 per 100,0000 citizens. He con-

cluded that living in a small-group society is significantly more dangerous than being a member of a more complex one.

Pinker uses Keeley's data and unpublished studies by economists to argue that complex society brought standing armies, laws, walled cities, and other innovations that restricted tribal fighting and protected the average citizen from violent crimes. "Hobbes understood this without having the data," Gat adds.

Pinker blames what he calls "anthropologists of peace" for distorting the record on small-scale group violence. "The classic 'gentle people' "—the Semang of the Malay peninsula, !Kung in Africa, and Central Arctic Inuit—"turned out to have higher homicide rates than those of American cities," Pinker says. He criticizes what he calls a single-minded determination "to make hunter-gatherers seem as peaceful as possible."

Such charges puzzle some biological anthropologists and archaeologists—the kinds of scholars who gather the type of data used in this debate. They do not argue for a Rousseauian perspective. But that doesn't mean they're ready to embrace a Hobbesian view, either. They find the data too weak to support such sweeping claims and add that the statistical averaging done by Pinker and Gat erases the enormous variation in small-scale societies. Pinker "misused the bioarchaeological record by selecting a few populations … biased toward supporting his argument," complains archaeologist Gwen Robbins Schug of Appalachian State University in Boone, North Carolina.

In a 2011 review of existing violence data on both nonstate and state societies, criminologist Amy Nivette of the University of Cambridge in the United Kingdom found persistent problems with data on small groups, which typically do not record violent deaths. Different researchers, for example, assign different rates of violence to the !Kung. And tiny numbers in small groups make statistics unreliable. Keeley cites the Polar Eskimo, for example, but given their small population, a single Eskimo murder every 50 years would equal the current rate for the United States. To sort out these problems, "researchers must carefully and more systematically consider what is meant by 'low' levels of violence," Nivette says. "Violence in nonstate societies is more complex and changeable than the stylized Hobbes-Rousseau dichotomy," she concludes.

Other researchers say they are more interested in the variability in violence than the averages. For example, Keeley's data do show dramatic differences in homicide rates among hunter-gatherers who tend to live in small bands and among agriculturalists in modest villages. More recent anthropological surveys by other researchers note that the Amba of Uganda and the Eastern Pueblo of New Mexico have fewer than two murders per 100,000 people, while the Hewa of New Guinea top 700 per 100,000, and the 19th century Kato of California reached 1450. By comparison, the U.S. national homicide rate peaked just below 10 per 100,000 during the crime-ridden late 1980s; in 2010, the cities of Baltimore and Detroit, which rank as the most violent, were nearly tied with just over 34 homicides per 100,000.

The great debate. Debra Martin and Gwen Robbins Schug argue that violence varies too widely for generalizations, while Azar Gat and Steven Pinker say civilization brought peace (*clockwise from upper right*).

Those living in small-group societies "are not always peaceful, and not always at one another's throat," says George Milner, an anthropologist at Pennsylvania State University, State College, who has spent much of his professional life examining ancient skeletons with signs of physical injuries. "We are beginning to see a highly varied picture—times when conflict is quite severe, as with today's Yanomamö [foragers of northern South America], and also places where there are prolonged periods of peace."

He cites the example of the Hopewell culture of the 1st through 5th centuries C.E. in eastern North America, which appears to have been "socially permeable," allowing traders to safely transport obsidian from sources in what is today Wyoming as far east as Ohio. Such ease of movement would have been unthinkable before and after that era, when violence between groups was more common. The interesting question, Milner says, is what changed. "To see this from a solely Hobbesian viewpoint misses the real story," he adds. "We want to know why people switch from peace to war and back again."

Pinker acknowledges that there are exceptions to the rule of more violent small-scale societies, but he maintains that "the bulk of the distribution includes massive death rates by violence." His point, he says, is that more organization leads to less chaos—and less violence. That extends to farming groups as well as hunter-gatherers, notes Gat, who cites the "staggeringly violent mortality rates" among agricultural groups in New Guinea.

Pinker and his critics agree that questions of how violence shifted in form, as well as rate, are worthy of study. To do that, scientists say they must gather more data on the ways of violence among a host of societies, both large and small, across the globe and through many millennia. Milner argues for tracking the pulses of conflict that characterize all societies to understand the conditions that spark war and pave the way to peace. But that goal will no doubt require fractious researchers from many disciplines to themselves lay down their arms and work more cooperatively. **–ANDREW LAWLER**

IN BATTLE

TWEETING THE LONDON RIOTS

IT STARTED WHEN FOUR GUNSHOTS RANG OUT ON FERRY Lane in north London on 4 August last year. Television and radio reported the known facts of the case: Police officers had killed a 29-year-old man, possibly unarmed, in an impoverished neighborhood. Online, rumors and rage spread rapidly. Some accused the police of executing the man. Two days later, London convulsed with the worst riots in living memory. As traditional media outlets scrambled to keep up, people used online social media tools to share information and opinions. Soon the riots infected other U.K. cities. For five nights, mobs clashed with police, smashed windows, looted shops, and started fires across the United Kingdom.

During the riots, Emma Tonkin watched and worried like everyone else. The streets had erupted in chaos in nearby Bristol, and she wondered whether her housemate should go to work the next day. "Why was this happening, the senseless violence?" she says. Tonkin, who is an information technology researcher at the University of Bath, studies online social media. According to press accounts, rioters were using media such as Twitter to incite and organize. That was a claim that she knew how to test. So she assembled a research team, harvested 600,000 tweets from the days of the riot, and got to work.

People have rioted since the dawn of civilization. But online social media may be radically changing how riots form and spread. Until recently, it was difficult for a large crowd of strangers to coordinate, be it for peaceful protests or violent riots, says Katharina Zweig, a network scientist at the University of Heidelberg in Germany: "You needed to contact a lot of people, make sure that nobody else could overhear your message." Now, with diffuse conversations of millions of people taking place on platforms like Facebook and Twitter, "you can direct messages to many other people at no cost."

Tonkin and her colleagues are part of a growing group of social scientists tracking online conversations, hoping to make sense of how they affect real-world events. For example, after the London mobs dispersed, *The Guardian* newspaper launched a collaboration with researchers from the London School of Economics and elsewhere to study whether social media had contributed to the magnitude of the riots, interviewing arrested looters, reviewing court records, and harvesting 2.6 million tweets. The stakes were high. Some had advocated giving the government the power to temporarily shut down social media sites.

Extracting meaning from social media is no easy task. In the case of Twitter, the enforced brevity gives rise to an abbreviated language that can baffle computer analysis. Another problem, Tonkin says, is that computers "have no sense of humor." Take, for example, "I am going to destroy that city," which sounds like terrorism to American ears but in Britain signals the intention to get drunk on holiday. "Our tools aren't good enough yet," she says, so researchers must still use human eyes to reality-check social-media data. And because Twitter users generate more than 200 million tweets per day, the crucial first step is to find what is relevant. Luckily, Twitter users organize their conversations with easily searchable hashtags, such as #londonriots.

Tonkin's first surprise was how visual those conversations were. About half of the tweets contained Web links, and the most popular led to photographs snapped by people in the street. One of the most iconic showed a sign on a London restaurant: "Due to the imminent collapse of society, we regret to announce we are closing at 6 p.m. tonight."

To explore whether rioters used Twitter to organize attacks, Tonkin's team tagged tweets by the sentiments they expressed, checking samples by eye to make sure the algorithms didn't stray far from reality. On the basis of that analysis, Tonkin declares the site innocent. "Clearly, Twitter did amplify signals," she says, acting as a megaphone both for those calling for violence and those calling for calm. But in her sample, inflammatory comments "were generally condemned" by those retweeting them. On the contrary, she says, the most popular use of Twitter for organizing was for cleanup crews.

The Guardian team also absolved Twitter, but they fingered less transparent online social media. At least two people attempted to use Facebook to set meeting points for mobs. And based on their interviews, the researchers concluded that many rioters were indeed using BlackBerry Messenger to coordinate attacks. The private network links BlackBerry phone users but creates no public trace on the Internet.

Of course, what was once private is increasingly shared online for all to see. Completely closed networks like BlackBerry Messenger may give way to Facebook, which is only superficially private. Governments and companies are already harvesting data from those online conversations. Could crowd phenomena such as riots one day be predicted, like weather? And by selectively shutting down social networks, could riots be controlled or even prevented? Possibly, Zweig says, but "I'm actually not sure we would want that as a society."

–JOHN BOHANNON

CIVILIZATION'S DOUBLE-EDGED SWORD

Recent archaeological finds from the Near East to Southeast Asia of ancient massacres raise questions about how violence changed as societies became more complex

THE CORPSES WERE LEFT TO ROT BY THE hundreds in the hot Syrian sun. The victors whittled three dozen or more of the human arm and leg bones into pointed sticks, perhaps as tools to further desecrate the skulls. After at least a few weeks, the decomposing remains were dragged a short distance and pushed into a trash pit. Then the victors butchered cows, sheep, and goats for an extravagant barbecue. When they were done, the celebrants hurled the picked-over animal bones and the ceramic plates, made especially for the occasion, onto the decaying heap. Then they returned to their homes in the nearby prosperous city.

As fighting engulfs the cities of modern Syria, the scene above is stark proof that mass warfare and civilization share a 6000-year history. The killing field at Tell Brak, about 500 kilometers northeast of Damascus near the Iraqi border, documents what is perhaps the world's oldest known large massacre or organized battle. When the violence occurred in about 3800 B.C.E., this settlement was evolving into one of the world's first fledgling cities.

What took place here and at other sites where complex societies were starting to coalesce is of particular interest to researchers studying human violence. Some scientists argue that civilization replaced tribal anarchy with a more organized way of life that reduced rates of violence (see p. 829). But recent finds around the world suggest an upsurge, not a decline, in violence during the key period when societies transitioned from the simpler organization of tribes and chiefdoms into complex urban life.

Even in the Near East, which has been intensively excavated for more than a century, scientists still grapple with fundamental questions about violence. There has been surprisingly little physical evidence for warfare, massacres, or even widespread murder here between the rise of agriculture around 10,000 B.C.E., when people were living in less complex groups, and the emergence of the Akkadian Empire around 2300 B.C.E. Recent finds in northern Syria, however, suggest that violence flared as urban life first began to take hold between 4000 B.C.E. and 3200 B.C.E.

Making their point. Tell Brak dig director Augusta McMahon examines a human bone shaped into a pointed tool.

The large settlement of Hamoukar was destroyed around 3500 B.C.E.; a team there found hundreds of what they maintain are sling bullets (*Science*, 31 August 2007, p. 1164). Even more startling, in 2007 and 2008, University of Cambridge archaeologists found three mass graves dating from about 3800 B.C.E. to 3600 B.C.E. at Tell Brak. The oldest and largest grave was at least 20 meters long and 4 meters wide, and included a jumbled pile of at least several hundred people—by far the earliest undisputed example of an event of mass violence. In a 2011 paper in the *Journal of Field Archaeology*, dig director Augusta McMahon and colleagues note that the majority of the dead were between the ages of 20 and 35. Although poor bone preservation makes it difficult to establish gender and cause of death, the state of the disarticulated bones suggests that the individuals all died at the same time and were left in the open for weeks or months. Many of the victims had previous head injuries that had healed.

Feast and famine. The rotting corpses of these Tell Brak victims were thrown into a pit with the remains of a massive barbecue.

The bodies' exposure, delay in burial, the careless collection of the rotting corpses—which ignored smaller bones from the hands and feet—and the casual disposal all "show an extraordinary disregard" for the victims, McMahon says, implying "that the dead were enemies." That is underscored by the fact that more than 40 human bones were whittled to make tools, which McMahon suggests were used "to deflesh and empty trophy skulls," given that some skulls have deep scratches. The apparent victors then celebrated an astonishing feast involving as many as 75 cattle and 300 sheep and goats. The remains of this repast were tossed on top of the corpses and then covered with dirt.

Whether the victims were locals or outsiders remains unclear. Study of tooth enamel from the bodies shows that at least some suffered from malnutrition. Tell Brak sits in a region prone to long-term drought and may have stored a surplus that drew hungry mobs; it also had workshops that harbored a wealth of beautiful objects that may have attracted envious invaders or spurred civil war.

McMahon suspects internal strife because Tell Brak was so populous in this period that it was not very vulnerable to outsiders. She notes that the callous treatment of corpses may have been "a useful control device" to deter internal discontent. Whatever the fight, it was clearly an organized killing field. "Here you see mass violence … motivated or controlled by central authorities," says archaeologist Henry Wright of the University of Michigan, Ann Arbor.

Peaceable kingdom?

On the other side of Asia, in the jungles of Thailand and Cambodia, complex society

CREDITS (TOP TO BOTTOM): AUGUSTA MCMAHON; MAHAN KALPA KHALSA

arrived several millennia later. Small villages predominated here for centuries, until about 900 C.E., when the first complex society arose around Angkor Wat in Cambodia.

Until recently, there has been a remarkable dearth of evidence for violence here during the region's Iron Age before the rise of Angkor. In that period, which began about 500 B.C.E., iron tools and weapons appeared, and tribes appear to have coalesced into more organized chiefdoms run by an elite.

But recent work shows that Southeast Asia was no peaceable kingdom just before or during the rise of Angkor. At a site in Cambodia called Phum Snay, about 80 kilometers northwest of what became the capital of Angkor Wat, archaeologists from Australia and New Zealand have found an array of sophisticated military artifacts such as swords (some more than 1 meter in length), daggers, spearheads, and epaulettes, as well as signs of violent conflict dating between 100 C.E. and 300 C.E.

Researchers estimate that of 30 skeletons from the town's cemetery, nearly a quarter had traumatic bone lesions, with males having more than females. Given the location and type of injury, these wounds are more likely due to interpersonal violence than accidents, says biological anthropologist Kathryn Dommett of James Cook University in Townsville, Australia, lead author of a 2011 *Antiquity* paper on the subject. And because soft-tissue wounds are not recorded, the level of violence likely was even higher.

This data meshes with the results from digs across the border in Thailand. Combining satellite imagery with fieldwork, archaeologists have mapped dozens of large settlements surrounded by elaborate moats and ramparts from this era. "This is a period of intense political development and possibly increased competition," says archaeologist Dougald O'Reilly of the Australian National University in Canberra, who led the Cambodian dig and has published on the Thai finds.

That is borne out at the Thai site of Noen U-Loke. Archaeologist Charles Higham of the University of Otago in Dunedin, New Zealand, found evidence for intercommunity conflict in a sudden proliferation of iron projectile points in the Iron Age, including one lodged in the spine of a young adult male; an older female had her head cut and smashed. Other sites, such as Ban Wang Hai, a village just to the north, have yielded iron swords and other weapons not associated with hunting.

O'Reilly says that the prevalence of "blunt force wounds" in this period points to either ritual warfare or a scramble to control resources such as iron. "There is a lot of evidence of increased warfare" just before the rise of Angkor, Higham says. O'Reilly believes that such conflict arose as populations increased along with competition for resources. As at Tell Brak, when other types of social complexity rose, so did the scale and complexity of warfare.

Archaeologist Glenn Schwartz of Johns Hopkins University in Baltimore, Maryland, notes that "early complex societies were able to organize much more effective killing machines, given their administrative and technological capabilities and large populations." So while laws, fortifications, and armies may have protected the bulk of the citizenry, when warfare did take place, it could be on a greater scale and ferocity than among the preceding smaller groups. "The nature of warfare changes," says anthropologist Patrick Nolan of the University of South Carolina, Columbia. "The frequency may decrease but the scale goes up."

Remains of the day. Murder victims at Phum Snay show traumatized wounds and are examined by O'Reilly (*inset, on left*) and a colleague.

CREDIT: KATE DOMETT

Historian Azar Gat of Tel Aviv University in Israel warns that the spectacle of large battles may mask the more important truth that a given individual in a complex society would be less likely to die from violence. "Ramses II took 20,000 troops to fight the Hittites," he says of the 13th century B.C.E. Egyptian king. "But the population of Egypt was 2 or 3 million. They were largely sheltered."

Even as they document cases of violence in early states, archaeologists are hesitant to generalize about long-term trends because archaeology can provide only glimpses of the past. For example, ancient texts describe bloody battles in the Near East in the 2nd millennium B.C.E., but archaeologists have found few sites to support the textual history. That's why single finds, such as the one at Tell Brak, can quickly rewrite old views—but should be considered cautiously, researchers say. "Serendipitous data discovery in archaeology frequently refutes statistical laws," says Yale University archaeologist Harvey Weiss.

Instead of trends in rates of violence, some researchers focus on how it assumed new forms, such as institutional slavery and human sacrifice, that are not seen in simpler societies. "Warfare and slavery go hand in hand" in the ancient world, Nolan says. Most slaves, he says, were captives of war, often the wives and children of slain soldiers. Brutal human sacrifice also appears as early states consolidate and display their power (see p. 834).

Many complex societies quickly developed moral codes and written laws designed to protect the young, the poor, and the defenseless. But they also found a galaxy of reasons to punish nonviolent behavior with violence, as U.S. sociologist Steven Spitzer and French philosopher Michel Foucault have noted. Sexual deviance, religious heresy, and betrayal of the state all could be punished with tortures and extended imprisonment undreamed of by our ancestors in simpler societies. The rise of civilization was indeed a double-edged sword.

–ANDREW LAWLER

Published by AAAS

Royal prerogative. An artist's impression of the death scene in a royal tomb at Ur from *The Illustrated London News* in 1928.

THE ULTIMATE SACRIFICE

Seeking to impress both gods and humans, early state societies across the globe displayed their power by ritually killing human victims

THE 63 SKELETONS WERE ARRANGED IN THE sealed death pit like actors on an eerie stage set. Just outside the King's Grave, archaeologists found six soldiers lined up, still wearing helmets and "guarding" the royal tomb. Beside them were two ox-drawn carts with drivers, grooms, and oxen lying nearby. Rows of men and women lined the passage to the tomb, and courtesans with elaborate golden headdresses sat in a circle around a set of musical instruments. This was a "theatre of public cruelty," enacted at the death of a Sumerian ruler about 4500 years ago in ancient Mesopotamia, according to an initial report by Leonard Woolley, the British archaeologist who excavated the royal tombs of Ur in Iraq in the 1920s and 1930s.

Woolley concluded in 1934 that these courtesans and servants had drunk some "deadly or soporific drug" from cups and a large copper cauldron he found in the pit. Most scholars accepted his account that the victims had gone willingly to their deaths, to serve their ruler in the netherworld.

So, when three researchers at the University of Pennsylvania (Penn) Museum of Archaeology and Anthropology in Philadelphia took their first look recently at computed tomography scans of two skulls from the death pit, they got a big surprise. "Holy cow!" said paleoanthropologist Janet Monge when she saw unmistakable radiating fractures from a blow to the side of a skull. This wasn't a case of mass suicide à la Jim Jones, but the ritual murder, or sacrifice, of 63 humans.

The Penn team proposed in a report in *Antiquity* last year that the retainers "were felled with a sharp instrument, heated, embalmed with mercury, dressed and [only then] laid ceremonially in rows." The ornaments and helmets had obscured the damage from the mortal blows for decades.

This new look at the victims of Ur is one of a flurry of multidisciplinary studies that has recently documented a macabre trail of human sacrifice that leads to every corner of the world, from the death pits of Ur and China to burials atop the highest peaks of the Andes.

Video featuring author Ann Gibbons.
www.scim.ag/sacrifice_vid

Using rigorous forensic and bioarchaeological methods, researchers have been able to reconstruct victims' last days and hours, and sometimes their identities, testing controversial claims of human sacrifice. "This is an exciting time for this kind of research," says biological anthropologist John Verano of Tulane University in New Orleans.

Researchers are finding that although human sacrifice was not frequent in most cultures, it was pervasive, taking place at one time or another in just about every ancient civilization in which someone had the rank and power to decide who died, Verano says. Although human sacrifice was seen as barbaric by classical times, it persisted in Rome, the Americas, and elsewhere until the rise of Judeo-Christian and Islamic religions that condemned it. Across cultures, most cases shared twin motivations: to please the gods, and to vividly assert and display rulers' power. For early states, whose rulers were consolidating power, ritual sacrifice seems to have been one way to discourage outside attacks and internal revolt by sowing fear. The cross-cultural data are beginning to give researchers an idea of "key patterns in the origins, motivation, and methods of [sacrifice]," says bioarchaeologist Haagen Klaus of Utah Valley University in Orem.

Myth or reality?

It doesn't take a scholar to guess from the friezes of Roman temples, images on Maya pots, or scenes in ancient Greek plays that our ancestors might have sacrificed one another. Historical accounts—from Herodotus in Greece and Pliny the Elder in Rome to Spanish priests in the Americas—recount sacrifices made by the Scythians, Etruscans, Romans, Incas, Aztecs, and Norse. Engraved labels on ancient Egyptian jars suggest that some early rulers took servants and concubines with them to the next world. Art on ceramic urns show a Maya god "sitting down to a plate of human hearts, just like a Maya king would eat a plate of tamales," says bioarchaeologist Andrew Scherer of Brown University.

Although depictions of ritualistic decapitation and dismemberment are found in the art and literature of many societies, convincing physical evidence has been rare until

CREDIT: UNIVERSITY OF PENNSYLVANIA MUSEUM

recently. "People didn't really look at marks of perimortem violence, so they didn't see the evidence," says bioarchaeologist Vera Tiesler Blos of the Autonomous University of Yucatán in Mexico.

This led some researchers to challenge the claims for human sacrifice in general and in Mesoamerica in particular. "I don't think that what we say is human sacrifice is anything other than [deaths in] war," says archaeologist Elizabeth Graham of University College London, who studies Maya sites in Belize. She notes that victims are often captives taken in war. "All societies have socially sanctioned killing," she says, citing the Holocaust of Germany as a particularly grievous recent example. "The poor Aztecs have been made out to be the most brutal people in the world, but if it's actually warfare, they killed few people."

Researchers agree that iconography and texts alone can't confirm sacrifice. For example, ethnographic accounts claim that the Aztecs slaughtered 80,000 war captives when dedicating the Great Pyramid of Tenochtitlan in 1487, but this is widely considered an exaggeration. However, in the past 15 years, researchers at Aztec sites have excavated sacrificial knives and stones, some with traces of human blood, as well as bones with cut marks and signs of heart extraction. This has "led us to conclude without a doubt that human sacrifice was a basic practice of Aztec religion," says Leonardo López Luján of the Templo Mayor Museum at the National Institute of Anthropology and History in Mexico City. At Tenochtitlan, his team found 47 decapitated bodies and 42 children with slit throats.

Death of a child. The Inca cut open this Muchik child's chest and removed the heart about 500 years ago.

So many new cases of sacrifice have been documented in the past decade that researchers classify them informally. There are retainer burials where slaves die with their owners; offerings of prized children; dedicatory burials that are a sort of bloody feng shui to bless buildings, such as Tenochtitlan during construction; and ritual killings of captives from war.

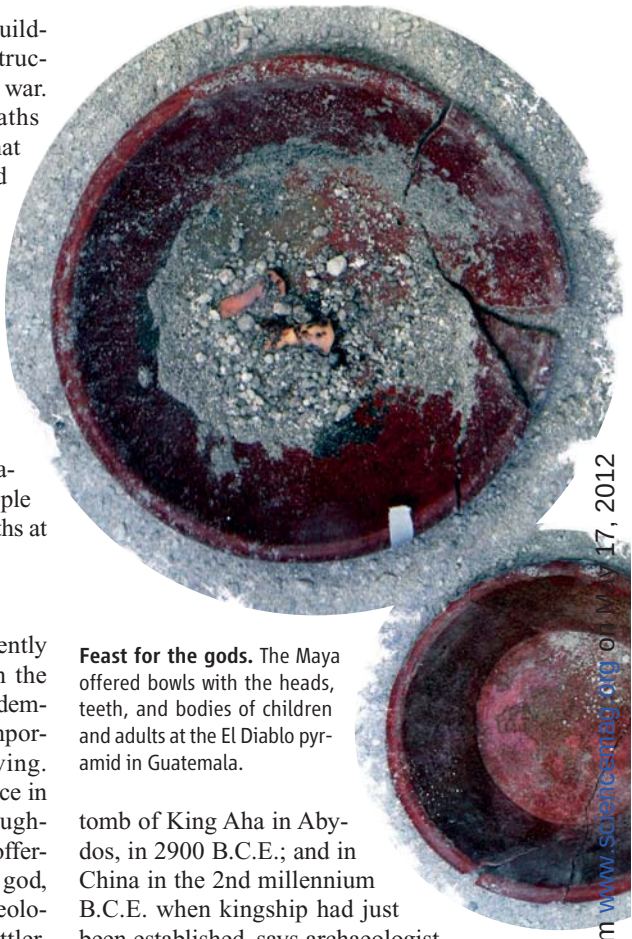
The difference between these deaths and other state-sanctioned killings is that sacrifice is ritualistic. Researchers add that they aren't targeting any particular society; indeed, a major finding is that human sacrifice was found in most emerging city-states around the world, particularly under a new ruler or in times of crisis. At the same time, it was relatively rare within populations. "Not everyone gets a sacrifice at their funeral," Scherer says. Klaus agrees: "There's not a lot of trauma in the populations at large. But a special subset of people did die extremely brutal and violent deaths at a variety of sites."

Loyal subjects

Retainer sacrifice, as at Ur, was apparently performed so that rulers could live in the afterlife much as they did in life, and to demonstrate their importance to the living. "It's not a sacrifice in the sense of slaughtering a cow or offering meat" to a god, says Penn archaeologist Richard Zettler. At Ur, the court attendants were set up as though they were at a banquet with food, drink, and music. They were adorned in golden wreaths studded with lapis lazuli and carnelian.

But did those who died really play the roles of guards, grooms, and courtesans in life? Strontium isotopes in bones and teeth show that two retainers at Ur were born locally and were not foreign captives, suggesting that they were indeed servants, says Penn archaeologist Aubrey Baadsgaard.

Such extravagant retainer sacrifices were rare. At Ur, the practice appears in only 16 out of about 2000 graves unearthed in the Royal Cemetery. But it also occurs in Egypt, at the tomb of King Aha in Abydos, in 2900 B.C.E.; and in China in the 2nd millennium B.C.E. when kingship had just been established, says archaeologist Glenn Schwartz of Johns Hopkins University in Baltimore, Maryland. "When you establish a new kingdom, a new kind of political organization with a ruler at the top, very often there is this strategy of making a big show of the power of this new social order by having this kind of retainer sacrifice," Schwartz says.



Feast for the gods. The Maya offered bowls with the heads, teeth, and bodies of children and adults at the El Diablo pyramid in Guatemala.

At Ur, the number of sacrificial victims and wealth of the treasures declines from about 2600 B.C.E. to 2450 B.C.E. The practice also declines and then vanishes in Egypt, perhaps because it was too costly to bury such wealth, both in objects and human life, or because established kings didn't need such a conspicuous display of power.

The Maya also practiced a form of retainer sacrifice in which some victims were children. In 2010, Brown University archaeologist Stephen Houston and his colleagues found six blood-red cache vessels beside a king's body in an airtight chamber of the El Diablo pyramid in the jungle near El Zotz, Guatemala. Interred in about 350 C.E., the caches contained the heads, teeth, and bodies of six children, aged 6 months to 5 years. The smallest were stuffed in the bowls whole, but the older children had been dismembered.

CREDITS: (TOP RIGHT) PROYECTO ARQUEOLÓGICO EL ZOTZ; (BOTTOM LEFT) SAM SCHOLES/ LAMBAYEQUE VALLEY BIOHISTORY PROJECT

Published by AAAS

Several had been ritually burned around the face and chest with low heat.

This matches previously known Maya iconography, showing children burning in large bowls with their hearts cut out, Scherer says. Such sacrifices "don't seem to have anything to do with warfare," Scherer says. "The Maya are replicating myths, with scenes of child sacrifice to the maize god."

Mountaintop maiden. This 15-year-old Inca girl was sacrificed atop Volcán Llullaillaco in Argentina 500 years ago.

Sacrificial lambs

Children were victims in other cultures, too, perhaps because they are often seen as the most precious offering. The Inca, for example, built platforms high in the southern Andes, where they held mountaintop ceremonies called capacocha, in which they sacrificed beautiful, unblemished children.

For example, a 15-year-old girl called the Llullaillaco Maiden was discovered in 1999 with a 7-year-old boy and a 6-year-old girl atop the 6739-meter-elevation Volcán Llullaillaco in northwest Argentina (http://scim.ag/LMaiden). The children were buried about 500 years ago, with gold and silver figurines. Two had headdresses of white feathers; one, a silver bracelet. Their youth and rich gifts suggest they were not captives of war, says archaeologist Johan Reinhard of the National Geographic Society in Washington, D.C.

The children were apparently treated well, consistent with ethnographic records suggesting that it was an honor to be chosen for this sacrifice. Stable isotopes from the Maiden's hair showed that her diet changed dramatically about a year before death, from a peasant's diet to one suddenly rich in meat and maize, an elite food; her diet shifted to more grains a few months before her death, as she trekked to the peak. "Children were specially selected and treated royally perhaps a year before they were taken up to the mountaintop," Verano says.

Child "sacrifice doesn't mean giving up those you don't like," Klaus says. "It's giving up those that matter the most." And the Inca may not have thought of their children as dying. "In this very sacred mountain environment, they'd be seen as living with the gods," Reinhard says. "They would, in essence, become deified."

And yet even Inca priests may have had an eye to impressing other humans as well as the gods, says bioarchaeologist Tiffiny Tung of Vanderbilt University in Nashville. The feting of the children en route to the peaks and the hubris of staging sacrifices at such lofty heights would have inspired awe and fear, helping the Inca assert power over their vast empire, Reinhard says.

Captive audience

There are many instances of reverential child sacrifice, but researchers agree that killing captives after battle may have been the most common kind of human sacrifice. Performed by cultures as diverse as the Aztec, the Wari, and the Shang Dynasty in China, this practice involves more than merely disposing of captives, those who study it insist. It is designed to shock and awe both enemies and subjects.

Trophy head. The Wari of Peru made a trophy of this captive foreigner's head in a highly ritualized process.

For example, in the 1980s, Chinese researchers uncovered 14 skulls in a row, including one placed inside a bronze food steamer, in the royal cemetery of Anyang, the capital of the ancient Shang dynasty in east-central China. The researchers assumed the skull fell into the pot by accident. Then in 1999, another skull turned up in a steamer in a tomb in a later Shang capital, according to archaeologist Tang Jigen of China's Academy of Social Sciences. This "leads us to the inescapable conclusion that the Shang people did indeed have the cruel custom of steaming human heads," he said in a recent publication.

Anyang fits the profile of cultures that sacrifice captives: At about 1200 B.C.E., it was the center of the country's first expansive power. Archaeologists found up to 15,000 sacrificial victims during digs in the 1930s and 1950s, and are now examining them in detail. Most are men of military age who were decapitated, Jigen says.

The men's arms and legs were frequently cut off in similar ways, suggesting they were killed ritually rather than in battle, and the human remains are mixed with animal bones. Few pits contain goods such as pottery that are included in typical burials. Shang oracle bones provide hints of sacrificial procedures: One inscription made on a defleshed skull mentions the decapitation of an enemy leader.

The deaths may mark the ritual killing of war captives in order to provide food or slaves to ancestors, says archaeologist Roderick Campbell of New York University, noting that the pits often contain remains of cattle, dogs, grain, wine, and other material commonly used in sacrifices. Later dynasties did not continue the penchant for sacrifice, about which later Chinese annals are silent.

Halfway around the world, the iconography of the Moche of northern Peru also suggests brutal sacrifice of war captives, done in ways that highlight the victor's power. Images show captives being paraded naked with bloody noses before a warrior priest and having their throats slit. But until the

CREDITS (TOP TO BOTTOM): JOHAN REINHARD; TIFFINY A. TUNG

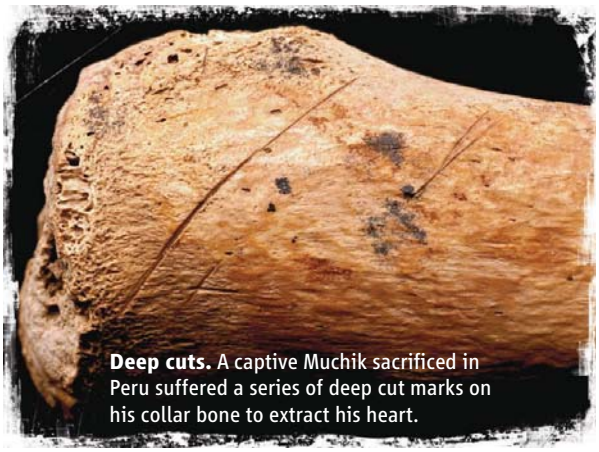
Baptism by fire. The Maya offered babies to their gods, as shown in this mythological scene.

1990s, some researchers thought that such scenes depicted Moche mythology or staged drama, not reality.

The overwhelming bioarchaeological evidence of hundreds of sacrificial victims, gathered since the 1990s, contradicts that view, Verano says. For example, the remains of more than 100 young men in a Moche plaza at the pyramid of Huaca de la Luna were either left exposed on the surface to be buried by windblown sand, or were incorporated in the fill of plazas during their construction around 500 C.E., Verano says. Analysis of the remains suggests these victims were captives brought back from battle, as they had wounds that had partially healed. Patterns of cut marks on the neck vertebrae and other bones confirm they were decapitated and that their bodies were defleshed. This was, Verano says, a "prominent display of military victory."

The study of sacrifice is also illuminating the politics and social structure of ancient societies. For example, researchers knew that after the Moche collapsed, their descendants, the Muchik, were ruled by another culture, the Sicán. Both cultures practiced sacrifice. But the details suggest that the Sicán governed loosely, Klaus says, because the Muchik still killed victims, often children, the traditional Moche way: The children's throats were slit, their chests were cut open to remove hearts, and their bodies buried with long-standing Moche funerary rituals. "It's a Moche template," not a Sicán one, Klaus says.

In other cases, researchers are using the existence of human sacrifice to show that certain cultures were more organized and sophisticated than had been realized. For example, a few scholars have suggested that the Wari of central Peru were not a state-level society. But Tung says their practices of human sacrifice, found as early as 600 C.E. to 1000 C.E. at Conchopata, suggest state-level control and organization.



Deep cuts. A captive Muchik sacrificed in Peru suffered a series of deep cut marks on his collar bone to extract his heart.

Tung and Kelly Knudson of Arizona State University, Tempe, have analyzed stable isotopes in 72 Wari trophy heads, many of children, and buried bodies. Of 29 properly buried bodies, all belonged to local people. But almost all of the trophy heads came from foreigners. This suggests that they were captives, according to a report last year in the *Journal of Anthropological Archaeology*.

Captives were brought alive to Conchopata, beheaded, and then processed into trophy heads in a "very systematic, very standardized way," Tung says. "They clearly had a standardized tool kit for drilling holes on the top of the head for the cord, so the heads would be upright and facing forward when displayed." This matches drawings on large ceramic urns, which show Wari warriors seizing prisoners and carrying trophy heads. "This is important, because it suggests you have Wari state structures used to promote this"—to coordinate the warriors, the priests who made the trophy heads, and the artists who depicted them on urns, Tung says.

The practice suggests a state-level society asserting its absolute authority against outsiders, Tung says. "Sacrifice is very orchestrated—it's not just death on the battlefields. It's a performance to demonstrate to your internal community and outsiders your absolute power."

As more cases of sacrifice emerge, some defy classification. This suggests that researchers have just begun to exhume the myriad ways that humans killed each other in the name of the gods and the state. "Our ability to see sacrifice in the past was somewhat limited. Now we're able to expand that view," Monge says. "I'd say we're just coming to realize in some measure the enormity of the violence of humans against humans."

–ANN GIBBONS

With reporting by Andrew Lawler.

CREDITS (TOP TO BOTTOM): JUSTIN KERR; SAM SCHOLES/LAMBAYEQUE VALLEY BIOHISTORY PROJECT

IN BATTLE

FIGHTING RITUALS

BLUSTER SAVES LIVES. CERTAIN AGRICULTURAL TRIBES in New Guinea are known for their "nothing fights," in which armed warriors gather by day to yell and gesture but rarely come close enough for combat. Such shows of force can go on for days and weeks, with men heading home at night and taking mutually agreed-upon breaks. It's one of many ways people manage conflict so that few get hurt and détente is maintained.

But it's not just humans who engage in ritual combat. Groups of birds known as green wood hoopoes engage in calling contests (see p. 828), and some species of ants settle their differences with mock tournaments. For example, honeypot ants of the U.S. Southwest stage elaborate, ritualistic face-offs against nearby nests, involving hundreds of ants and sometimes lasting for days. "Displays are such an important way of conveying information because it's

a way of avoiding the risk of physical injury," explains Andrew Radford, a behavioral ecologist at the University of Bristol in the United Kingdom. Group displays convey the group's ability to compete. That birds and ants mount these shows of force suggests a common evolutionary pressure to minimize actual losses and demonstrates that such behaviors can emerge without organized leaders or higher thought processes.

Ants, which make war on each other and will fight to the death to defend nests and territories, are "the most aggressive animals," says behavioral biologist Bert Hölldobler of Arizona State University, Tempe, who has studied these insects for decades. Yet the honeypot ants (*Myrmecocystus mimicus*) stage fights that don't kill. These ants, which are each less than a centimeter long, eat termites, dead insects, and honeydew, and earned their name because some colony members store food in their swollen abdomens, becoming living "honey pots." The ants' foraging territories often overlap with those of nearby nests.

In a ritual contest, two honeypot ants meet head-on and stiffen their legs as though on stilts. They may climb up onto a pebble, as though trying to appear larger. They hold their abdomens and their heads high and probe each other with their antennae. Ants from the same colony back off within a few seconds, whereas those from different colonies turn sideways to one another, lift the abdomen even higher, and point it at the opponent. They kick and drum their antennae on each other for up to 30 seconds, then disengage, uninjured, and move on to the next ant.

Typically, the smaller ant yields to the larger ant; analyses of films of tournaments suggest that the ants try to fool one another into "thinking" they are larger, says Hölldobler. A single foreign ant can trigger the displays, which then grow as scouts recruit up to several hundred workers to the tournament.

During tournaments, the ants appear to track encounters to get a sense of which side has more numbers, says Hölldobler. The recruiting scouts do this, never engaging an opponent for more than a few seconds and wandering throughout the tournament area.

In many ways, the displaying ants are the equivalent of ritualized antler butting in male deer. These bouts reduce the chances that rivals will be hurt or killed in disputes. That makes sense from an evolutionary perspective, because in a real fight, even the stronger opponent risks injury. For example, in another species of ant, Australia's meat ant (*Iridomyrmex purpureus*), displaying ants line the boundaries of each colony's territory and help maintain a stable border at a minimal cost to the colony, says evolutionary biologist Ellen van Wilgenburg of the University of Melbourne in Australia. "If most confrontations between non-nestmates involved potentially fatal fighting, then the conflict zones would act as a significant drain on the workforce," she notes. In 2005, she and her colleagues showed that if these displays escalated into real fights, resident ants tended to be the aggressor, most likely because they had more to lose.

In the honeypot ant contests, if one group outnumbers the other, the tournament shifts ever closer to the smaller group's nest, interfering ever more with foraging by its workers. If the contest is uneven enough, the smaller group retreats and closes off the nest entrance, hiding out until the rivals have gone back to their own nests. Sometimes the retreat isn't fast enough, and the larger group raids the nest of the smaller one, carrying off the young, stealing the honey pots, and killing or running off the queens.

Video featuring ants in tournaments.
http://scim.ag/ants_vid

The same breakdown can happen in human "nothing fights." As University of Oxford archaeologist Barry Cunliffe wrote in 2006 in the book *Conflict*, if one side fails to show up, a rout can ensue. And the losers wind up abandoning their villages and sometimes being killed.

–ELIZABETH PENNISI

Dovish females? Many scholars contend that women are not inherently more peaceful. Margaret Thatcher led the United Kingdom into war.

GENDER AND VIOLENCE

Researchers are probing links between the status of women in a society and its propensity toward war

WHEN ERIK MELANDER FIRST ENCOUNTERED a series of papers linking gender inequality and war, the premise struck him as fishy. Melander had trained as a conflict scholar in the 1990s, when the forces underlying war were relatively well established: a lack of democracy, a low level of economic development, and the presence of nationalism. The status of women wasn't even on the list.

So he was not convinced when, in 2000, he read the first in a group of studies by political scientist Mary Caprioli of the University of Minnesota, Duluth, that challenged some closely held views about violence and war by connecting the low position of women to conflicts from international aggression to civil war. The idea that the status of women helped predict a state's volatility sounded "like wishful thinking," recalls Melander, deputy director of the Uppsala Conflict Data Program at Uppsala University in Sweden. Part of the problem was what he assumed was a simplistic approach to sex differences underlying the connection. The notion that women are biologically so hard-wired for peace that simply giving them more say in international affairs yielded tranquility seemed suspect. "I expected that when controlling for other factors," he says, "any relationship would go away, or the effects of gender equality would be very small."

Then he tested the hypothesis himself. Melander checked the claim that gender inequality correlated with escalated levels of conflict within states. Measuring the status of women by looking at the sex of a country's highest leader, the proportion of women in the legislature, and the ratio of women to men who receive higher education, he controlled for factors like democracy, economic development, and the time since a country's last civil war. To his surprise, the results, published in *International Studies Quarterly* in 2005, confirmed the finding that had aroused his suspicion. In a second study published that same year, he broadened the picture:

Risk factor. A high proportion of teen mothers can suggest low status for women, which correlates with violence and war.

States where women were oppressed also had higher rates of political imprisonments, killings, and disappearances.

Melander's research is among a nascent body of work in international relations showing gender inequality to be an important security barometer. By focusing on gender inequality rather than biological sex differences, these researchers say they have identified a previously overlooked trigger of conflict. Causality is far from proven, however, and some critics say that gender inequality could be a proxy for other underlying causes.

Caprioli, whose work first piqued Melander's interest, is in some ways the ringleader of the new group. From 2000 to 2006, she published a series of widely cited statistical analyses linking the low status of women to a host of negative phenomena. After controlling for other factors, she found states where women are treated poorly are more likely to become embroiled in disputes with other states, more likely to turn to violence in those disputes, and more likely to erupt into civil war.

Still, she recalls a reviewer writing on one of her first published papers: "I don't recognize this as research." She says: "It was very hard in the beginning because I was pioneering a new field of study."

Born warriors?

Throughout human history, males have been the more violent sex. Skeletons unearthed from early human societies show more head injuries among males, and men are believed to have been the aggressors as well as the victims. Gender clearly matters in conflict. But sorting out just how and why it matters—and how significant a role biology and culture each play—has proven a thorny task.

One of the first works to tackle that challenge was the 2001 book *War and Gender*, by political scientist and American University professor emeritus Joshua S. Goldstein. "That was a seminal piece of work," says Ismene Gizelis, a political scientist at the University of Essex in the United Kingdom. "It brought the issue of gender into international relations and conflict studies."

The book grew out of a puzzle. If both

CREDITS (TOP TO BOTTOM): © JEAN LOUIS ATLAN/SYGMA/CORBIS; © BILL GENTILE/CORBIS

gender roles and war practices vary widely from one culture to the next, Goldstein wondered, why was there not more variation when it came to gender roles in war? When he dug into research from biology, ethology, and anthropology, the obvious explanation—that biology turns women into doves and men into hawks—didn't completely explain the conundrum. To be sure, Goldstein determined that there are some core biological differences that affect war behavior. Childbirth and motherhood, for example, keep women away from the battlefield. But he contends that inborn traits are not as influential as commonly portrayed—and that biological factors interact with cultural ones.

Testosterone levels, for example, can spike following shifts in a man's social status, such as getting married or winning a game. "Cultures are responsible for the exaggerated gender roles we see in societies," Goldstein says. Melander agrees that although "there are evolutionary roots to the male warrior role," a strictly biological view fails to "take into account the differences between, say, Sweden and Pakistan."

If a propensity toward fighting were inborn, men might disproportionately report more enthusiasm for war. But even as Goldstein was researching *War and Gender*, studies were showing that men and women are often remarkably similar in their views on war. In the 1990s, for example, political scientist Mark Tessler and demographer Ina Warriner, then at the University of Wisconsin, Milwaukee, looked at attitudes toward conflict in the Middle East. They relied on surveys from Israel, Egypt, Palestine, and Kuwait in which participants had been asked questions such as, "Do you believe that the Arab-Israeli conflict can be solved by diplomacy or is a military solution required?" They found that men and women barely differed in their answers.

But when Tessler and Warriner shifted their focus away from an individual's gender toward his or her perspective on gender equality, the picture changed. Along with questions about international aggression, the Middle Eastern survey participants answered questions such as, "Do you think it is more important for a boy to go to school than a girl?" Comparing the responses, the scholars found a strong association between sexism and bellicosity. "You can predict how a person stands on war and peace and on the Arab-Israeli issue based on what they think about gender equality," Tessler says.

Conflict scholars who focus on gender inequality believe values governing how states behave abroad reflect values within a society. "If you respect women, you also respect the rights of others," Gizelis says. "As a result, you also deal with conflict in a different way." Political scientist and University of Maryland, College Park, professor emeritus Ted Robert Gurr has found the same to be true of ethnic discrimination: Societies in which minorities suffer widespread discrimination are also more volatile.

When war does break out, moreover, equal societies tend to be better at restoring peace. In 2009, Gizelis examined 124 civil wars from 1945 to 2000. Evaluating the status of women using indicators such as female-to-male life expectancy ratio and secondary school enrollment ratio, she found that U.N. peacekeeping operations were far more likely to succeed in states where men and women were relatively equal before the war.

Essential ABCs. War-torn societies where girls leave school prematurely are less likely to restore peace.

Some now hold up this body of work as evidence that raising the status of women is not just a good in itself: It can also help ensure global security. That idea is percolating through policy circles. In a 2005 speech on the empowerment of women and girls, U.N. Secretary-General Kofi Annan said, "I would venture that no policy is more important in preventing conflict, or in achieving reconciliation after a conflict has ended."

Caprioli is more reserved: "This link is but a piece of the puzzle," she says. But, she adds, if promoting gender equality "decreases war by any percent, I would argue it is a change well worth making."

Others warn that correlation does not equal causation. Goldstein believes there is a link between gender inequality and war but says, "The big question is the direction of causality. Does gender inequality make societies more war-prone, or does being involved in lots of warfare exaggerate gender inequality?" Women, indeed, are among the most affected victims of war.

Yale University political scientist Nicholas Sambanis cautions that the relationship might be explained away by a third factor. He ventures that the level of gender inequality might be "a proxy for other, more fundamental things, like cultural differences, rule of law, [and] institutional development."

No one has yet proven the existence of such a third factor. And Melander counters that causality is also difficult to show with other, more established explanations of conflict, such as a low level of economic development or an undemocratic regime.

Better data may eventually help scholars unravel the connection between gender inequality and war. The World Economic Forum's *Global Gender Gap Report* and the U.N.'s Women's Indicators and Statistics Database both compile statistics on the status of women, such as the proportion of female professional and technical workers in a country and the percentage of women who have given birth by age 20. But because many of those statistics were introduced only in the last decade, scholars comparing conflict across time are limited to indicators for which good longitudinal data exist—and there are problems with all of them. The presence of a female leader can be misleading, because many women rise to power through dynasties—think Indira Gandhi or Benazir Bhutto. And while a low fertility rate, an indicator used by Caprioli in several papers, often suggests a higher status for women, there are notable exceptions.

In the meantime, conflict scholars who are focused on gender inequality say the largest issue is being taken seriously. Some investigations of war looking at dozens of variables disregard gender inequality entirely. "Gender-related approaches still are not well integrated into the mainstream of conflict studies," Goldstein says. "It's more like two camps living under the same roof but barely talking." Melander contends that if those ignoring the research were to instead test the hypothesis, as he did, they might change their minds.

–MARA HVISTENDAHL

IN BATTLE

FROM WAR TO PEACE

KISS AND MAKE UP. ANIMALS FROM DUNG BEETLES TO chimpanzees fight each other, but the ability to reconcile differences after a spat was once considered a uniquely human trait. Then in the 1970s while observing chimpanzees at a zoo in the Netherlands, ethologist Frans de Waal witnessed a male attacking a female. Immediately, other males came to her rescue, and soon afterward, amid hooting by the troop, the male and female embraced. In 1979, he published a paper documenting that after a fight, a whole lot of kissing, embracing, and holding hands went on between chimp opponents. De Waal, now at Emory University in Atlanta, called this behavior reconciliation, and researchers have since observed it in more than 30 primate species.

Like people, animals that live and work together do better if they get along. And that means smoothing over differences, particularly between individuals who have long-standing, valuable relationships. "In the last 30 years, we have learned a lot about conflict management in primates," says Filippo Aureli, now at the University of Veracruz in Mexico. "It is important to verify if animals other than primates behave in a similar way." Finding these behaviors in other parts of the animal kingdom sheds light on their evolution and cognitive basis.

Researchers first turned their attention from chimps to other mammals. A 1998 study found that goats that have tussled over food later nuzzle and groom each other. Hostile dolphins rub each other gently with their flippers or tow one another by a flipper to reestablish alliances after a fight. In 2010, Alessandro Cozzi of the Phérosynthèse Research Institute in Semiochemistry and Applied Ethology in Saint Saturnin Les Apt, France, showed that horses, too, make up after aggressive encounters by grooming, sniffing, playing with, or following one another. Studies in canines and hyenas have also found evidence of postconflict management strategies.

Nicola Clayton, an ornithologist at the University of Cambridge in the United Kingdom, and her colleagues wanted to look beyond mammals. Birds and mammals "have very different evolutionary histories," notes Orlaith Fraser, a cognitive biologist at the University of Vienna. If they shared these traits, that would suggest convergent evolution and might point to the circumstances under which these strategies evolve.

In 2007, Clayton's team studied postconflict behavior in a group of 10 captive rooks, highly social, intelligent birds related to crows and ravens. They watched the free-flying but captive birds in their large cage each afternoon and took note of when one displaced another or showed other signs of aggression. For the next 10 minutes, they recorded whether either the opponents or another bird made any peaceful gestures, say, touching beaks or sharing food. They were looking not only for reconciliation but also for "consolation," actions by an uninvolved bird to soothe the victim.

They saw no evidence of reconciliation, which they interpreted as a sign that the relationships between opponents weren't very important. But they did find that the mates of victimized birds would try to console the loser, which makes sense, as rooks pair for life early on and have a lot invested in each other.

Meanwhile, Thomas Bugnyar, now at the University of Vienna, had been following up on a pilot study in another species of bird, ravens, which suggested that these birds do smooth ruffled feathers after fights. In 2004, he monitored postconflict activity in a group of hand-raised ravens. Unlike rooks, young ravens can spend several years single, during which time they may develop friendships with other birds that can help them fight for dominance and access to food. Among these birds, he and Fraser often saw victims sidling up to bystanders or, after a particularly long chase or contentious encounter, bystanders offering consolation to losers. Typically, the bystanders were "friends" of the victim—they had hung out together, groomed each other, and shared information and food—Fraser and Bugnyar reported 12 May 2010 in *PLoS ONE*. They saw hints of reconciliation, but nothing significant.

However, when they took a close look at the data from the pilot study of seven ravens studied in 2002, Fraser and Bugnyar did find evidence that opponents made up after fighting—if they had been in cahoots beforehand. "The key seems to be that there are strong social relationships," Bugnyar says. As he and Fraser reported 25 March 2011 in *PLoS ONE*, reconciliation in these cases was worth the risk that a second fight might break out with renewed contact. Indeed, reconciliation reduced subsequent aggression. "The results are very similar to the results you get in primates," Bugnyar says.

Preliminary studies in ravens in a natural setting suggest that what happens in captivity happens in the wild, as well. "These behaviors are more widespread than we thought," Fraser says. Like primates, the birds "have similar needs and have the same solutions to the same problems."

–ELIZABETH PENNISI

HUMANCONFLICT
http://scim.ag/hum_conflict

DRONE WARS

Are remotely piloted aircraft changing the nature of war?

New flight plan. The U.S. Air Force now trains more pilots for crewless aircraft like the MQ-1 Predator.

THE VIDEO IS SLIGHTLY GRAINY, BUT IT'S EASY enough to make out a white car parked on a dirt road and a person nearby setting up what looks like a tube. The video, shot in Afghanistan, comes from a remotely piloted aircraft (RPA), or drone, that's been following two men. Good intelligence suggests they are planning a mortar attack on a coalition airbase. The man in the video drops something into the tube and ducks for cover. Smoke shoots from the tube as a round is fired. The video jumps ahead to the man and his accomplice driving away. Small crosshairs superimposed on the video never waver from the car as it follows a riverside road. Suddenly there's a bright flash and a cloud of black smoke, the impact of a laser-guided missile launched from the drone.

What happens next is just as important, according to Colonel James Bitzes, a legal adviser to the U.S. Air Force who showed the video at a conference last year at the New America Foundation, a think tank in Washington, D.C. (Bitzes's presentation is online at http://bit.ly/NAFBitzes; the drone strike video begins around 38:20). People gather around the smoldering car, and some of them begin tossing objects from the trunk into the river. Another car pulls up and someone dumps an object from the trunk of that car, too. An Army team later recovered the objects—weapons—from the river, Bitzes said. "With the benefit of the RPAs, … there's a possibility of gathering additional intelligence," he said.

Given the potential to track down bad guys and kill them without risking American lives, it's no wonder the U.S. Department of Defense plans to spend $30.8 billion on developing and acquiring RPAs between 2011 and 2015. Dozens of other countries have or are pursuing drones of their own. Thanks to improvements in artificial intelligence, these machines will become more capable of making decisions, including, perhaps, whether to kill humans. As the technology zooms forward, experts are scrambling to catch up with the psychological, ethical, legal, and policy implications for 21st century conflict.

Killing at a distance

Not everyone agrees that RPAs represent a fundamental shift in military technology. "You see people breathlessly saying this is a revolution in warfare, but I think that's overstating it," says Werner Dahm, chief scientific officer for the Air Force between 2008 and 2010 and current director of the Security & Defense Systems Initiative at Arizona State University, Tempe. "There's been a continuing progression in technology ever since the stone-throwing days. I think it's more like the sling is to the hand-thrown stone." Dahm notes that the U.S. Air Force and intelligence agencies have been developing drones since World War II.

All in a day's work. Waging war from halfway across the globe, RPA operators may not face the same combat stress as deployed troops.

But unlike most of their predecessors, today's drones are operated half a world away from the war zone. "It's making the decisions even more remote from the action," says Peter Hancock, a human factors researcher at the University of Central Florida in Orlando. "We really don't know what all the ramifications of that are."

There's little doubt drones are changing the experience of combat for those who operate them. Operators work at U.S. bases in dark, cool rooms designed more for protecting electronics than for ensuring human comfort, and they communicate with troops on the ground and intelligence analysts across the United States via Internet chat rooms. "Most of the time, I get to fight the war and go home and see the wife and kids at night," one operator told Peter Singer, who directs the 21st Century Defense Initiative at the Brookings Institution in Washington, D.C. The quote comes from Singer's 2009 book *Wired for War*.

As Singer and others have pointed out, drones raise questions about the psychology of killing. On one hand, the physical distance between the person pulling the trigger (or pushing the button) and the target has never been greater. In his influential 1995 book, *On Killing*, former Army Ranger Lieutenant Colonel Dave Grossman argues that killing at a distance is easier, psychologically, than killing at close range. Drawing on firsthand accounts gleaned from historical records and his own interviews with soldiers, Grossman writes of a human resistance to killing that increases with proximity: "This process culminates at the close end of the spectrum, when the resistance to bayoneting or stabbing becomes tremendously intense, and killing with the bare hands … becomes almost unthinkable."

CREDITS (TOP TO BOTTOM): AIR FORCE; DAMIAN DOVARGANES/AP PHOTO

On the other hand, the drones' cameras and other sensors may close the psychological distance. Operators track targets for hours or even days. They watch targets do bad things, but they also see them go home and play with their kids, says Colonel Hernando Ortega, a flight surgeon at the Air Force Intelligence, Surveillance and Reconnaissance Agency at Lackland Air Force Base in Texas. "It can be very personal," Ortega says.

The psychological impact of killing in combat—from a distance or otherwise—is a neglected and somewhat taboo area of research, says Shira Maguen, a psychologist at the San Francisco Veterans Affairs Medical Center in California. Since 2009, Maguen and her colleagues have published a series of studies of veterans of the Vietnam, Gulf, and Iraq wars that found that soldiers who report killing in battle suffer more symptoms of post-traumatic stress disorder (PTSD) than do those who saw similar levels of combat but did not report killing. Alcohol abuse and issues with anger and violent behavior also appear to be more common in those who've killed. Maguen says her work with deployed troops suggests that those who kill in self-defense seem to cope better. "With the drone pilots, they're not experiencing any personal threat, and I don't think we understand how that's important."

One of the first efforts to examine the psychological health of drone operators was led by Colonel Kent McDonald and Wayne Chappelle, who head the neuropsychiatry and aerospace psychology departments, respectively, at the Air Force School of Aerospace Medicine at Wright-Patterson Air Force Base in Ohio. In surveys given to 874 RPA operators and 628 airmen in other logistics and support work at the same bases, they found higher levels of stress and job-related fatigue in the RPA operators, the vast majority of whom worked on unarmed RPAs used for surveillance. The main causes of stress were long hours, rotating shifts, and the nature of the job, which entails hours of monotony punctuated by brief flurries of intense activity.

Only 4% of drone operators in McDonald and Chappelle's study screened positive for heightened risk of PTSD. That's far lower than the 12% to 17% estimated prevalence of PTSD in veterans of the Afghanistan and Iraq wars, and comparable to the prevalence of PTSD in the general population. Only a small subset of these RPA operators participated in combat missions, and McDonald and Chappelle plan to investigate whether combat experience puts operators at greater risk. Ortega says he's not surprised by these findings. He attributes the apparently low risk of PTSD in RPA operators to a combination of training that mentally toughens all service members and the fact that even RPA operators on combat missions don't experience the same fight-or-flight activation of the sympathetic nervous system as soldiers on the battlefield.

Drone diversity. Smaller RPAs like the RQ-11 Raven (*top*) can be deployed in the field, while the MQ-9 Reaper (*bottom*) can cover long distances.

CREDITS (TOP TO BOTTOM): JUSTIN KERR; SAM SCHOLES/LAMBAYEQUE VALLEY BIOHISTORY PROJECT

Robot arms control

Removing the combatant from the war zone carries ethical and legal implications, says Armin Krishnan, a political scientist specializing in international security at the University of Texas, El Paso. "Targeted killing [of specific individuals] is becoming easier with drones," Krishnan says, because it reduces the risk of American casualties, and with it, the political risk of launching an attack. Krishnan sees the covert U.S. drone attacks as particularly problematic. (U.S. officials acknowledged conducting a drone campaign in Pakistan only recently, but it was long an open secret: The New America Foundation estimates that 296 drone strikes have killed between 1785 and 2771 people in Pakistan since 2004, the vast majority of them militants.) Little is known about how the Central Intellgence Agency determines who is a legitimate target, Krishnan says: "The question is, is killing always justified? There's no public accountability for that." He and others worry that drones make it easier to go to war. In a January article in *The New York Times*, Brookings's Singer noted that Congress has never debated the drone campaign in Pakistan. "What troubles me … is how a new technology is short-circuiting the decision-making process for what used to be the most important choice a democracy could make," he wrote.

Another debate swirls around technical advances that are making weapons systems more autonomous. The touchiest issue is whether a machine should ever make the decision to kill. Ronald Arkin, a roboticist at the Georgia Institute of Technology in Atlanta, argues that in some cases, a robot might make better ethical decisions than a human soldier blinded by anger, frustration, or fear. As a proof of concept, he and colleagues developed a prototype "ethical governor" that could be built into an autonomous weapons system. This software would kick in once the system locks onto a target and is ready to fire, and it would abort any attack that violates international laws of war or U.S.-specific rules of engagement.

Another program they developed, a "responsibility adviser," would ensure that a human is ultimately responsible for every action executed by the system. Arkin stresses that he hopes autonomous weapons will never be used, but given the history of warfare, he thinks it's an inevitability that must be prepared for.

Determining whether a given attack violates international law isn't nearly as hard as discriminating combatants from civilians and other protected individuals, says Noel Sharkey, an expert on artificial intelligence and robotics at the University of Sheffield in the United Kingdom. Unfortunately, Sharkey says, "we have nothing like this at the moment." He and Arkin disagree about how long it will take to develop. "Ron says maybe in 20 years; I say more like 100 or so," Sharkey says. In 2009, Sharkey and three academic colleagues founded the International Committee for Robot Arms Control, which hopes to promote international discussion, and ultimately a ban, on autonomous weapons. "If the technology is developed, I'm pretty certain it will be used, ready or not," Sharkey says. "That's my real worry." **–GREG MILLER**

Published by AAAS

HUMANCONFLICT
http://scim.ag/hum_conflict

REVIEW

ANCESTRAL HIERARCHY AND CONFLICT

Christopher Boehm

Ancestral *Pan*, the shared predecessor of humans, bonobos, and chimpanzees, lived in social dominance hierarchies that created conflict through individual and coalitional competition. This ancestor had male and female mediators, but individuals often reconciled independently. An evolutionary trajectory is traced from this ancestor to extant hunter-gatherers, whose coalitional behavior results in suppressed dominance and competition, except in mate competition. A territorial ancestral *Pan* would not have engaged in intensive warfare if we consider bonobo behavior, but modern human foragers have the potential for full-scale war. Although hunter-gatherers are able to resolve conflicts preemptively, they also use mechanisms, such as truces and peace pacts, to mitigate conflict when the costs become too high. Today, humans retain the genetic underpinnings of both conflict and conflict management; thus, we retain the potential for both war and peace.

Ancestral *Pan* was the shared antecedent of humans and our two genetically closest relatives, *Pan troglodytes* (chimpanzees) and *Pan paniscus* (bonobos), and, if we look for strong social similarities shared by the two living *Pan* species and human foragers, ancestral traits are readily identified. All three live in bounded social groups and fight with conspecifics (*1*), and all three have territorial tendencies (*2*, *3*), along with a substantial amount of dyadic dominance-and-submission behavior that can erupt into serious conflict countered by active, sometimes highly effective, peacemaking (*4*). In addition, all three form community-wide coalitions that cooperatively threaten males of other groups (*5*), whereas within their communities sizable coalitions of subordinate individuals may band together to reduce the domination of higher-ranking males (*6*). Here, I rely on a behavioral phylogenetic approach (Box 1) (*7*) that allows me to conclude that such shared traits are primitive and were to be found in ancestral *Pan*. By analyzing similarities across all three descendants of ancestral *Pan*, I can make conclusions about behaviors likely to have been present in our ancestors. From this estimate of our ancestral behavior, I can explore the factors that may have led to the more uniquely human set of behaviors we find in modern *Homo sapiens*.

Comparative Hierarchies

Although all three species have social dominance behavior, there are differences worth noting (*8*). As the least aggressive, bonobos (both wild and captive) have a rather complicated social organization. Not only do they exhibit both alpha females and males, but the only small coalitions involving males occur when mothers support their sons; the alpha male stays in place because his alpha mother supports him (*9*). In contrast, small coalitions of female bonobos readily form to compete with the larger males, and as a result females generally are codominant with males (*10*). For example, subordinate coalitions often form when two or more allied females routinely (and usually quite successfully) gain primary access to food resources at the expense of competing individual males. Such behavior can escalate to the point where a sizable female coalition may wound or even possibly kill a male (*11*, *12*). Bonobos are known for having conflicts that are far less frequent and less dangerous than those of chimpanzees and more bellicose human foragers, but social dominance hierarchies are clearly present. So although they resolve many of their conflicts dyadically through sexual behavior (*10*), triadic alpha-male power interventions are also reported (*13*), and fighting does occur.

Wild chimpanzees differ in several important respects. They also have linear hierarchies among adult males, who compete strongly for their power positions; however, female hierarchies are much looser. Because females in the wild seldom form coalitions, even lower-ranking males usually can dominate the strongest females (*14*). Chimpanzee males also engage frequently in coalitions with other males, forming competitive political alliances that further their individual quests for power within the male hierarchy. However, when a pair of rivals manages to unseat an incumbent alpha, only one of the two will assume the alpha position as a new set of competitive alliances comes into play (*3*, *15*). As with bonobos, among wild chimpanzees there are small and sizable subordinate coalitions that behave antihierarchically. For example, occasionally a large coalition of males, or males and females, will actively attack and temporarily "banish" or kill an adult male (*16*–*19*).

In some captive chimpanzee groups, the balance of power between males and females is modified in the bonobo direction because the females form strong bonds and act as cohesive subordinate coalitions. In this way, they are able to challenge, manipulate, and sometimes temporarily dominate high-ranking males. However, in both wild and captive conditions, powerful alpha male chimpanzees intervene frequently in dyadic conflicts as dominant peacemakers that coerce the protagonists to separate. Individuals will also reconcile on their own, and in close conditions of captivity female third parties also manipulate male conflicts without use of force, as though encouraging the protagonists to make up (*20*).

The contemporary humans most appropriate for this evolutionary analysis are mobile hunter-gatherers who continue the "culturally modern" behaviors associated with anatomically modern

Fig. 1. Depicted by the artist is an Australian Aborigine band under attack by members of another band. There was considerable violent conflict among bands in precontact Aboriginal Australia, but there were methods of effecting truces and peace treaties as well.

Department of Biological Sciences and Anthropology, University of Southern California, Los Angeles, CA 90089, USA. E-mail: cboehm1@msn.com

humans starting about 45,000 years ago (*21*) in the Late Pleistocene. Many scholars (*22*, *23*) have arbitrarily written off these potential exemplars because in recent times they have been so politically "marginalized"; however, I suggest that Late Pleistocene populations were similarly marginalized because of climate shifts. Thus, contemporary foragers—but just those who are economically and culturally independent and isolated from domestication and modern commerce—in fact are likely exemplars for what culturally modern humans were doing in the Late Pleistocene (*24*).

These "Late Pleistocene–appropriate" foragers live in small, multifamily egalitarian bands, and, although they certainly compete for mates and food, they are militant about not allowing male competition to develop into the kind of general dominance that is sought through alpha-male behavior in the two *Pan* species (*4*). This is because the type of antihierarchical subordinate coalition that is found in bonobos and chimpanzees emerges much more definitively in human foragers, where it results in significant reproductive leveling (*6*).

In these foragers, the virtual elimination of alpha-male behavior is possible because rebellious political coalitions are able to act efficiently on a moral basis, with a conscience-based sense of right and wrong as a political catalyst that intensifies negative group reactions against personal self-aggrandizement. In addition, both conflict and group-sponsored conflict management are vigilantly mediated by such moral concerns, which cannot be readily discerned in either chimpanzees or bonobos and therefore are likely not ancestral but derived (*24*).

In all three extant species, group males react hostilely toward males of neighboring groups, and in chimpanzees and humans lethal conflict behavior can result. However, bonobos mainly issue vocalized threats, and, although one case of possible wounding is reported (*13*), sometimes two groups may mingle (*9*). Regarding territoriality, bonobos diverge from the human-chimpanzee pattern, suggesting that intergroup conflict in ancestral *Pan* may not have been lethal often—even though chimpanzees and some human foragers systematically kill outsiders (*25*–*27*). By 45,000 years ago, however, humans had become "symbolic" and otherwise culturally modern (*21*), and surely our evolved potential for lethal conflict between bands had become substantial (*28*–*31*). Fortunately, we also had evolved ways of negotiating such conflicts when they became too costly (*4*).

Conflict and Its Resolution

In ancestral *Pan*, competitive dyadic interactions over food or mates did not necessarily provoke conflict—as long as clear roles of dominance and submission prevailed. The subordinate could simply flee, as an act of avoidance, or could stay in place and offer a submissive signal. In either case, fighting with a superior was avoided. However, because dominance hierarchies operated over time, individuals inevitably gained or lost power according to their stage of life, health, or coalitionary situation. Ancestrally, this made for circumstances in which, at times, subordinates were prone to challenge their superiors.

The result was conflict in the form of bluffing or fighting, and in fact, just like its three living descendants, this ancestor would have been well evolved to fight. As has been seen, it also was evolved to cope with conflict in that when serious quarrels erupted, if a higher-ranking individual was present, pacifying interventions could be initiated by an altruistic peacemaker (*13*, *15*, *32*–*34*). Furthermore, often the two parties might be capable of effective reconciliation (*20*) on their own. Thus, this ancestor possessed primitive patterns of conflict and conflict management that appear to have coevolved continuously for over 5 million years, as evidenced by their presence in the three descendant lineages (*35*).

Considering just the well-studied chimpanzee, conflict resolution appears to be cognitively sophisticated in that a variety of tactics are used by third parties in breaking up fights and making sure they are not resumed (*32*). In both chimpanzees and bonobos (*36*), there is also variety in the ways they reconcile their differences dyadically (*20*). Because this is also the case with humans (*33*), a flexible peacemaking approach is likely to be ancestral. Furthermore, captive chimpanzees show something like "community concern," in that female coalitions support males that act as peacemakers (*20*, *37*). Such behavior also is obvious in humans, but it is more difficult to demonstrate for bonobos, whose coalitions are not so large.

Ancestrally, it is likely that an additional type of conflict arose when subordinate coalitions formed to threaten or attack powerful, overbearing individuals that no individual group member would dare to attack alone (*4*). The resulting conflict could result in the disliked dominator fleeing or being wounded or killed. An intriguing aspect of such attacks was that they may not necessarily have been about food or mates; in all likelihood, this ancestor would have been disposed to fight over political position or power (*38*).

When such coalitional conflicts arose between allied subordinates and the alpha individuals who

Box 1. THE PARSIMONY PRINCIPLE IN BIOLOGY

One way to illuminate patterns of conflict present today is to examine their deeper background. At present, fortunately, ancestral hierarchies and the conflict they generated can be recreated with considerable reliability because such reconstructions can be based on the evolutionary biologist's specialized principle of **parsimony** (*61*, *62*).

Parsimony theory posits that over time natural selection is likely to operate as simply as possible, and the theory applies more strongly if the evolutionary time span between an ancestor and its living descendants is short. Thus, when today's humans, bonobos, and chimpanzees share a behavioral trait like conflict management, we can infer that the recent ancestor shared by all three of them at 5 to 7 million years ago (*63*) passed that trait down to each of them directly and therefore that the contemporary behaviors will have a similar underlying genetic basis. For instance, if the above three descendant species share conflict management, it is highly unlikely that their shared ancestor originally lacked this ability and that it was later acquired by all three independently.

Parsimony can be relied upon unless persuasive contrary evidence exists. Consider the fact that physically the above three species are only moderately sexually dimorphic with respect to body size. Parsimony would suggest that the same was true of their precursor, which I call ancestral *Pan* (*57*). This assumption has been challenged by an analysis of postcranial skeletal fossils for *Australopithecus afarensis*, which suggests the presence of dimorphism comparable to gorillas (*64*). This challenge has its problems, because there is no definitive evidence that *afarensis* was a member of the direct human lineage (*65*). In fact, *afarensis* may well have been a hominin descendant of ancestral *Pan* that went extinct. In reconstructing behavioral features of ancestral *Pan* that relate to conflict, it must be kept in mind that such challenges may arise and that in each case the facts must be weighed.

If a primitive trait is established ancestrally on the basis of parsimony, it follows that over time the trait would have been continuously present in the lineages that connected the three descendant species with their shared ancestor. However, the really definitive fossil referents go back only 2 million years; they include ourselves, archaic *H. sapiens*, and *H. erectus* (*21*). Although earlier fossil species include a number of candidates scholars espouse for human ancestry, such as *Ardipithecus ramidus* (*65*), ongoing debates leave even the relatively recent, tool-using *H.* (or *Australopithecus*) *habilis* as a mere possibility (*66*).

Two main rules for reconstruction help to keep the analysis conservative. First, if a major behavior pattern is found unanimously and unambiguously in bonobos, chimpanzees, and human foragers and if no evolutionary evidence exists to the contrary, the probability of its having been present in their (quite recent) shared ancestor will be very high (*1*, *3*). Second, I suggest that, if among the three extant species a trait is expressed strongly by two but weakly by the other, the ancestral assessment should be based just on the least common denominator (*24*).

normally dominated them, there were likely no pacifying interventions by third parties because this is absent today. However, overwhelming subordinate power would have made a quick and decisive end to such conflicts. These subordinate rebellions may well have provided the evolutionary basis for the egalitarian social orders that typify the life-styles of culturally modern hunter-gatherers and that in fact have continued to prevail among most tribal agriculturalists (*39*) [except when the advent of chiefdoms and civilizations has made for the rise of hierarchy (*23*)].

Thinking more broadly about sociopolitical basics of human nature, we are a species that is given both to within-group competition and to cooperation. Our innate competitiveness, coupled with the ability to fight, can lead to violence; however, we seem genetically predisposed to reduce stressful and often dangerous conflict levels. This can be accomplished by the parties reconciling, having a responsible third party intervene, or, sometimes, simply by spatial avoidance.

Human Foragers and Conflict

There is little doubt that conflict and conflict management have coevolved. Being competitive certainly has reproductive payoffs, but a capacity to end conflicts also is beneficial. It is this combination that defines much of today's political life.

Over the past 5 to 7 million years, humans have diverged from their two *Pan* congeners in several major respects that impinge on conflict and its management. First, at the level of the phenotype, we temporarily lost the alpha male role by becoming politically egalitarian (*40*). This means that we lost both a selfishly efficient oppressor and a forceful, but altruistic, peacemaker. Second, at the level of genotype, we acquired a conscience (with a sense of shame) that made us moral. This changed the very nature of our group life (*24*), for now, in addition to primitive, fearfully submissive reactions to the power of others, moral hunter-gatherers follow rules simply because group values support them. It seems we have evolved to internalize such values (*41*, *42*).

This thinking applies to all humans, but here we focus on how conflict and conflict management work in the simpler foraging bands we have been considering as later paleoanthropological exemplars. Today's evolutionarily appropriate foragers are of the type who are spatially mobile and highly cooperative and who vigilantly keep their egalitarian orders in place with only muted leadership. Because there are no alpha males to intervene authoritatively in their disputes, a serious dyadic conflict can quickly result in homicide. Indeed, the homicide rate per capita for egalitarian foragers is as high as in large American cities (*5*, *40*, *43*).

Within the community, evidence for "homicide" in adult chimpanzees and bonobos is mostly inferential but highly suggestive. For example, at Gombe alpha-male Goblin would likely have been killed by solo challenger Wilkie had not a veterinarian intervened (*16*), whereas at the Mahale field site the alpha male was photographically documented as being killed by other males (*18*). Among bonobos, a savage attack by half a dozen united females may have killed an adult male (*11*). Thus, ancestrally within-group conflict likely had at least some modest effect on adult mortality.

Aside from the important issue of morality as a derived behavior that intensifies social control and makes it more effective, in the area of conflict there are several other significant differences between humans and the two other species in our small clade. One is weapons. Bonobos and especially chimpanzees may use tools, but the use of weapons as humans do, to hunt sizable mammals, is totally absent (*44*). Bonobos and chimpanzees do have the potential to kill a smaller mammal, mainly using their canines (*45*), and this is also true of conspecific group attacks (*11*, *14*), which usually take at least several minutes for severe damage to be rendered. Human foragers use efficient hunting weapons to kill sizable mammals and members of their own species alike, and with these weapons they can do so much more quickly, at a distance, and often from ambush (*46*). These differences escalated the consequences of human conflict. Further escalation stemmed from the uniquely human propensity to lethally retaliate for the death of a close relative (*47*), a behavior that in all likelihood is not ancestral but which figures prominently in hunter-gatherer conflict. Thus, for humans the scope and consequences of serious conflict within the group would appear to be considerably greater than with ancestral *Pan*.

Another human difference is the understanding of death. When omnivorous chimpanzees or bonobos hunt, unlike dedicated carnivores, they have no evolved response that makes them into efficient automatic killers; in fact, prey may be eaten alive (*14*). When chimpanzee patrols savagely attack strangers they leave them battered and torn (*25*), but sometimes alive with some very small chance of recovery (*14*). This also is true of the one observed serious within-group attack by bonobos (*11*). In contrast, in spite of their diverse supernatural beliefs human foragers understand death as a termination of social responsiveness and muscular activity, and they inflict it deliberately. For instance, when egalitarian hunter-gatherers use capital punishment to eliminate despots, they shoot to kill (*24*).

Humans readily become lethal revenge-seekers, and chimpanzees and bonobos may at least try to retaliate for a prior aggression (*10*, *14*), so there were likely some modest ancestral preadaptations for such behavior (*47*). However, understanding how to kill with lethal weapons can lead to such motives becoming costly to groups, particularly when revenge becomes moralized as a matter of honor. On the other hand, being vindictive can be useful to a group if such a reputation keeps it from being attacked (*48*).

This holds for foragers that are given to conflict and even more so for clannish patrilocal tribal farmers (*49*, *50*). Among simpler hunter-gatherers, when a male kills another male, usually over a female, close relatives will predictably seek lethal retaliation (*40*), and the killer's only recourse is to move away. But with those foragers who do develop active, intensive raiding and warfare patterns, revenge needs also can help to motivate much larger attacks by entire groups (*51*).

Warfare is a major problem for modern humans, and most theories of warfare focus directly on resource competition (*52*). However, materialistic theories fail to fully explain the warfare patterns of forager societies (*31*, *53*). For instance, the Iñupiaq hunter-gatherers of northwest Alaska compete with some of their close neighbors for nearby natural resources, but at long distance they also conduct prolonged nonterritorial genocidal warfare against enemy bands, with surprise attacks and pitched battles motivated by retaliation (*51*). Here, I believe it is not necessary to favor one cause. A serious intergroup conflict may begin because of either resource competition or revenge, and the pattern can continue because of either factor, or both (*47*).

Managing Human Conflicts

Because adult male foragers are constantly armed as intentional killers of prey, any male dispute within a band can have drastic social consequences—particularly because our species is so prone to retaliation. On the other hand, with our large brains and language we can understand, and symbolically evaluate, the negative effects of disputes. Moral values play a direct role when band members not only condemn acts that create conflict, including killing, bullying, theft, and cheating (*54*), but extol the virtues of social harmony (*55*) and actively promote altruistic generosity that fosters cooperation (*56*). As a practical application, sharing large carcasses is so well regulated culturally that serious conflict over precious meat is routinely obviated (*57*). However, a noteworthy area that is poorly regulated socially, and which produces most of the serious conflict, is male competition over females (*40*).

Curiously, often conflictive behavior, per se, is not condemned morally. Rather, foragers appear to view most serious personal disputes as requiring an application of third-party mediation—but not individual punishment. However, it is clear that when a band suppresses morally deviant behaviors that are likely to create conflict, in effect it is preemptively reducing conflict levels as well (*58*). In evolutionary contexts this preemptive effect of social control has been little noted, but for leaderless hunter-gatherers it is very important precisely because forceful alpha peacemakers are lacking (*24*).

When dyadic conflicts do arise, powerless band members will try actively to distract the participants, facilitate negotiations, or otherwise serve as mediators (*33*). When disputes become serious, third-party intervention quickly becomes impotent, but spatial avoidance is often used by humans (*40*) if changing bands is an option. However, in

the Holocene, when many humans settled down as farmers, avoidance became too costly. This led egalitarian farmers, in their sedentary tribes, to allot some modest authority to their chosen leaders as conflict mediators. Eventually, far more authority was given to hereditary leaders as certain sedentary foragers or farmers lost their egalitarian ethos and began to form social classes. For farmers, this opened the way for the evolution of chiefdoms, kingdoms, and states, and in this way the protracted age of widespread egalitarianism ended (*23*).

The role of hereditary chiefs included an ability to mediate disputes with considerable authority (*59*); beginning about 5000 years ago, with the rise of centralized states, political leaders also had increasing coercive force to back up their mediating decisions in the form of a standing army paid for by taxation (*23*). This centralized approach to conflict management continues today in large modern societies, be they democratic or autocratic, which by their size and nature also require efficient centralized command and control. The job of managing potential major conflicts may be delegated largely to legal systems with police and courts of law, as well as to political figures who serve as mediators. But, in nations, a standing army can intervene if an internecine conflict becomes too heated. Such power interventions are not always successful, for civil wars are not uncommon. Indeed, the rebellious roots of "terrorism" would appear to be ancient.

With respect to conflict between groups, a conservative assessment of ancestral *Pan*'s patterns would limit this to the bonobo level of territorial behavior, based chiefly on threats (*13*), although a more chimpanzee-oriented model would suggest otherwise (*3*). If we look at today's humans, there has been a tendency for many hunting bands to fight ferociously with their foraging neighbors (Fig. 1) as single or regional units (*28*, *30*, *51*), whereas the sedentary egalitarian farmers who followed them are still more prone to raiding and warfare (*23*). In both cases, revenge motives can exacerbate these patterns, whereas intergroup conflicts sometimes can be negotiated through peacemaking or truces.

Virtually all nations devote significant resources to warfare readiness. In today's world, there seems to be a consistent, sizable number of small wars ongoing despite the fact that we have not experienced an active, truly global conflict since the Second World War (*59*). However, whereas a hunting band often can move away and simply avoid a predatory neighbor, our world of nations is firmly situated in space, and this ancient remedy cannot be applied (*60*). Fortunately, peace negotiations like truces and treaties that are found among warlike hunter-gatherers (*4*, *51*) also are used by nations. It would appear that, as derived behaviors, both violent, two-way intergroup conflict (*27*) and intergroup peacemaking have prevailed at least since humans became culturally modern.

Thus, humans are a conflictive species that developed its mechanisms for intergroup conflict management in the context of small hunting bands and did so without any formal political centralization. As our populations have grown, on a cultural basis we have rather quickly evolved into a species given to centralized governments. These governments try and often succeed in protecting their populaces from internal conflict and often are able to negotiate conflicts that arise between nations. This same command and control also enables us to mobilize efficiently for aggressive conventional warfare and to create weapons of mass destruction.

In the foreseeable future, the human capacity for political problem-solving will continue to be tested, with an ancient capacity for conflict management providing an important tool in international politics. The future of our own species (and surely many others) will be hanging in the balance.

References and Notes

1. R. W. Wrangham, in *The Evolution of Human Behavior: Primate Models*, W. G. Kinzey, Ed. (SUNY Press, Albany, NY, 1987), pp. 51–71.
2. M. P. Ghiglieri, *J. Hum. Evol.* **16**, 319 (1987).
3. R. W. Wrangham, D. Peterson, *Demonic Males: Apes and the Origins of Human Violence* (Houghton Mifflin, New York, 1996).
4. C. Boehm, *Hierarchy in the Forest: The Evolution of Egalitarian Behavior* (Harvard Univ. Press, Cambridge, MA, 1999).
5. R. W. Wrangham, *Yearb. Phys. Anthropol.* **42**, 1 (1999).
6. C. Boehm, *Am. Nat.* **150** (suppl. 1), S100 (1997).
7. S. Brosnan, *Soc. Justice Res.* **19**, 153 (2006).
8. C. B. Stanford, *Curr. Anthropol.* **39**, 399 (1998).
9. T. Furuichi, *Evol. Anthropol.* **20**, 131 (2011).
10. F. B. M. de Waal, F. Lanting, *Bonobo: The Forgotten Ape* (Univ. of California Press, Berkeley, CA, 1997).
11. I. Parker, *New Yorker*, 30 July 2007, p. 48.
12. Parker's journalistic account appeared in the *New Yorker*, but it was based on an interview with a long-term scientific bonobo field researcher.
13. T. Kano, *The Last Ape: Pygmy Chimpanzee Behavior and Ecology* (Stanford Univ. Press, Stanford, CA, 1992).
14. J. Goodall, *The Chimpanzees of Gombe: Patterns of Behavior* (Belknap, Cambridge, MA, 1986).
15. F. B. M. de Waal, *Chimpanzee Politics: Power and Sex Among Apes* (Harper and Row, New York, 1982).
16. J. Goodall, in *Human Origins*, vol. 1 of *Topics in Primatology*, T. Nishida, W. C. McGrew, P. Marler, M. Pickford, F. B. M. de Waal, Eds. (Univ. of Tokyo Press, Tokyo, 1992), pp. 131–142.
17. T. Nishida, *Pan Afr. News* **3**, 4 (1996).
18. S. Ladd, K. Maloney, Chimp murder in Mahale, www.nomad-tanzania.com/blogs/greystoke-mahale/murder-in-mahale (2011).
19. In this case (*18*), a tourist group documented the alpha male's being killed photographically, in the absence of scientific observers.
20. F. B. M. de Waal, *Good Natured: The Origins of Right and Wrong in Humans and Other Animals* (Harvard Univ. Press, Cambridge, MA, 1996).
21. R. G. Klein, *The Human Career: Human Biological and Cultural Origins* (Univ. of Chicago Press, Chicago, 1999).
22. R. L. Kelly, *The Foraging Spectrum: Diversity in Hunter-Gatherer Lifeways* (Smithsonian Institution Press, Washington, DC, 1995).
23. E. R. Service, *Origin of the State and Civilization: The Process of Cultural Evolution* (Norton, New York, 1975).
24. C. Boehm, *Moral Origins: The Evolution of Altruism, Virtue, and Shame* (Basic, New York, 2012).
25. M. N. Muller, J. C. Mitani, *Adv. Stud. Behav.* **35**, 275 (2005).
26. K. F. Otterbein, in *Handbook of Social and Cultural Anthropology*, J. J. Honigmann, Ed. (Rand McNally, Chicago, 1974), pp. 923–958.
27. R. W. Wrangham, L. Glowacki, *Hum. Nat.* **23**, 5 (2012).
28. S. Bowles, *Science* **314**, 1569 (2006).
29. S. Bowles, *Science* **324**, 1293 (2009).
30. L. H. Keeley, *War Before Civilization* (Oxford Univ. Press, New York, 1996).
31. R. C. Kelly, *Warless Societies and the Evolution of War* (Univ. of Michigan Press, Ann Arbor, 2000).
32. C. Boehm, in *Chimpanzee Cultures*, R. W. Wrangham, W. C. McGrew, F. B. M. de Waal, P. G. Heltne, Eds. (Harvard Univ. Press, Cambridge, MA, 1994).
33. D. P. Fry, in *Natural Conflict Resolution*, F. Aureli, F. B. M. de Waal, Eds. (Univ. of California Press, Berkeley, 2000), pp. 334–351.
34. R. B. Lee, *The! Kung san: Men, Women, and Work in a Foraging Society* (Cambridge Univ. Press, Cambridge, 1979).
35. Again, such reconstructions do not necessarily apply to side branches that have gone extinct, such as Neandertals or various other species that may not have been in the direct human line. In this context, there is some doubt as to whether even *H. habilis* was a direct ancestor of the modern humans (*21*).
36. F. B. M. de Waal, *Science* **289**, 586 (2000).
37. J. C. Flack, F. B. M. de Waal, *J. Conscious. Stud.* **7**, 1 (2000).
38. C. Boehm, J. Flack, in *The Social Psychology of Power*, A. Guinote, Ed. (Guilford, New York, 2010), pp. 46–86.
39. C. Boehm, *Curr. Anthropol.* **34**, 227 (1993).
40. B. M. Knauft, *Curr. Anthropol.* **32**, 391 (1991).
41. H. Gintis, *J. Theor. Biol.* **220**, 407 (2003).
42. H. A. Simon, *Science* **250**, 1665 (1990).
43. R. W. Wrangham, M. L. Wilson, M. N. Muller, *Primates* **47**, 14 (2006).
44. W. C. McGrew, *The Cultured Chimpanzee: Reflections on Cultural Primatology* (Cambridge Univ. Press, Cambridge, 2004).
45. C. B. Stanford, *The Hunting Apes: Meat Eating and the Origins of Human Behavior* (Princeton Univ. Press, Princeton, NJ, 1999).
46. J. Woodburn, *Man (London)* **17**, 431 (1982).
47. C. Boehm, *Br. J. Criminol.* **51**, 518 (2011).
48. C. Boehm, *Blood Revenge: The Enactment and Management of Conflict in Montenegro and Other Tribal Societies* (Univ. of Pennsylvania Press, Philadelphia, 1986).
49. N. A. Chagnon, *Science* **239**, 985 (1988).
50. K. F. Otterbein, C. S. Otterbein, *Am. Anthropol.* **67**, 1470 (1965).
51. E. S. Burch, *Alliance and Conflict: The World System of the Iñupiaq Eskimos* (Univ. of Nebraska Press, Lincoln, NE, 2005).
52. M. Harris, *The Rise of Anthropological Theory: A History of Theories of Culture* (Thomas Crowell, New York, 1968).
53. N. Chagnon, *Yanomamo: The Fierce People* (Holt, Rinehart, and Winston, New York, 1983).
54. C. Boehm, *J. Conscious. Stud.* **7**, 79 (2000).
55. E. Sober, D. S. Wilson, *Unto Others: The Evolution and Psychology of Unselfish Behavior* (Harvard Univ. Press, Cambridge, MA, 1998).
56. C. Boehm, *Cross-Cultural Res.* **42**, 319 (2008).
57. C. Boehm, *J. Bioecon.* **6**, 109 (2004).
58. C. Boehm, *J. Soc. Biol. Struct.* **5**, 413 (1982).
59. S. Pinker, *The Better Angels of Our Nature: Why Violence Has Declined* (Viking, New York, 2011).
60. C. Boehm, in *Evolutionary Psychology and Violence: A Primer for Policymakers and Public Policy Advocates*, R. W. Bloom, N. Dess, Eds. (Praeger, London, 2003), pp. 203–237.
61. T. J. Ord, E. P. Martins, in *Evolutionary Behavioral Ecology*, D. F. Westneat, C. W. Fox, Eds. (Oxford Univ. Press, New York, 2010), pp. 108–130.
62. E. Sober, *Reconstructing the Past: Parsimony, Evolution, and Inference* (MIT Press, Cambridge, MA, 1988).
63. M. Ruvolo, T. R. Disotell, M. W. Allard, W. M. Brown, R. L. Honeycutt, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 1570 (1991).
64. J. M. Plavcan, *Hum. Nat.* **23**, 45 (2012).
65. C. O. Lovejoy, *Science* **326**, 74 (2009).
66. A. Gibbons, *Science* **332**, 1370 (2011).

Acknowledgments: Relevant research support was forthcoming from the Templeton Foundation and the Goodall Research Center at University of Southern California; comments are acknowledged from H. Gintis, M. Muller, S. Bowles, and R. Wrangham.

10.1126/science.1219961

HUMANCONFLICT
http://scim.ag/hum_conflict

REVIEW

THE GROUP SELF

Naomi Ellemers

Although people often tend to consider themselves and others as unique individuals, there are many situations in which they think, feel, and act primarily as group members. This can bring out the best in them, as when they are inspired to help fellow citizens in need, or the worst, as when they show hostility against others simply because they represent another religious or ethnic group. Understanding when and why the group self becomes more important than the individual self, and how this affects people's thoughts, feelings, and behaviors, can help to prevent and redirect unwelcome aspects of human behavior by addressing them at the appropriate level of self.

The expression "group self" may appear to be a contradiction in terms. It often seems that people who come together in groups or crowds lose their sense of self. Even upright citizens can show a lack of self-awareness and self-control when put together in a group. Societal problems are often explained in this way. These include the aggression of soccer hooligans, the maltreatment of Iraqi prisoners by U.S. soldiers, or financial risk-taking by bankers. At the same time, groups—and the ensuing "loss of self"—can also bring out the best in people. Concern for society can help suppress selfish impulses. A focus on shared goals can inspire and encourage people to overcome personal doubts and anxieties that would otherwise prevent them from pursuing desired outcomes.

The group self is that part of people's self-view that is based on the groups to which they belong (Fig. 1). Rather than implying a loss of self, this indicates a (temporary) transformation of the conception of self from an individual to a group level, at which collective concerns become more important than individual differences (*1*). Understanding when and how people are driven by concerns relevant to their group helps explain important societal outcomes such as inter-religious or ethnic tensions, as well as individual outcomes, such as physical and mental health hazards deriving from membership in groups that are devalued in society (*2*).

Traditional approaches conceive of group behavior as stemming from individual goals, motives, and abilities. These assume that people work together with others because their insights, efforts, or resources can help them achieve attractive goals. They argue that people help each other because this indirectly benefits themselves, because of genetic overlap in families or because of ethnic and cultural similarities. They propose that people are friendly to others because they have agreeable personalities or are socially intelligent, or because they like those particular others. This focus on individual-level dispositions or concerns to explain group behavior is found across different disciplines in science such as law, philosophy, and ethics, but also in biology, sociology, work and organization sciences, and most prominently in psychology (*3*).

In the 1970s, Henri Tajfel and John Turner (*4*) took a different approach. Driven by the desire to understand social discrimination, aggression, and conflict between different groups in society, they developed what later became known as social identity theory. The theory specifies basic cognitive processes and social conditions that help understand and predict when and why people may think, feel, and act as group members and when they are more likely to respond as separate individuals. This seminal work (*5*) has informed analyses of a variety of group issues ranging from participation in political movements to leadership in organizations (*6*).

The willingness to act in line with group values or to invest effort in the achievement of group goals can shift from one situation to the next (*7*). The group self (instead of the individual self) can guide people's behaviors when they realize they are part of the group (cognitive self-definition) or when they are subjectively committed to the group (emotional self-involvement) (*8*). These two do not necessarily go together. The realization that one is part of a group raises concern about the group's fate or image. When abroad, it is easy to feel ashamed of the way fellow citizens of one's home country behave on their holidays. The image of the group also reflects on individual group members, even when they don't particularly care for the group. Feelings of commitment motivate people to display sol-

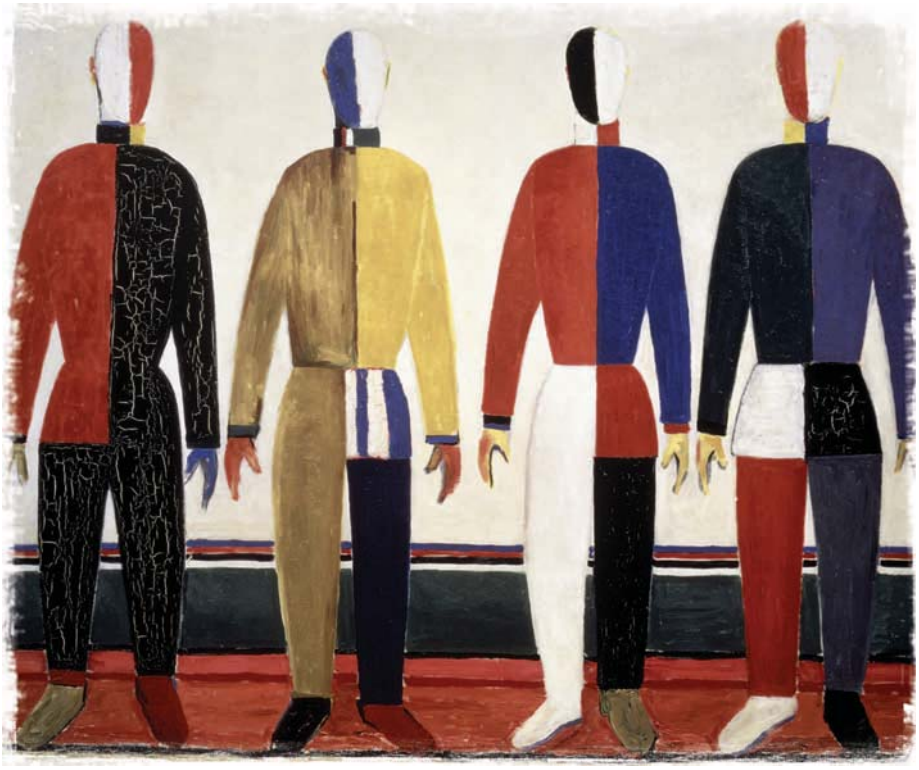


Fig. 1. *Sportsmen*, by Kazimir Malevich (dated 1928–1932): This painting depicts four sportsmen who are visually distinct from each other in their colors of dress. Yet they are not articulated as separate individuals with unique facial features, because their common identity as sportsmen is more important here.

Institute for Psychological Research, Department of Social and Organizational Psychology, Leiden University, Post Office Box 9555, 2300 RB Leiden, Netherlands. E-mail: Ellemers@fsw.leidenuniv.nl

idarity toward the group and its goals. This can happen when Peace Corps workers stand up for the rights of groups in need. They invest effort in the group because they care, even if they are not (fully) included in the group whose interests they protect.

Making Sense of the Situation

In theory, anyone can be considered as part of a number of different groups. These include groups people are born into (defined by gender, ethnicity, or nationality) as well as self-chosen affiliations, for instance with a particular professional, religious, or political group. These groups play an important role in how people see themselves and others (Fig. 2, row a), because they help to quickly make sense of a new situation (*9*). Deciding which group affiliations are situationally relevant (social categorization) helps people understand how others relate to them (social comparison) and determines which distinct attributes and behavioral norms are most relevant for them (*10*).

For instance, a study with preschool boys and girls showed they were equally able to remember the details of a video in which different children and adults chose and wore colored Mickey Mouse caps. After they had been asked to distinguish between activities children like (play with toys, eat candy) and activities that adults like (clean things, drink wine), they imitated the behavior of their own age group, regardless of the gender of the model. After having distinguished between activities that boys like and activities that girls like, they imitated the behavior of their own gender group model, regardless of age (*11*). Thus, when making their own choice, they took into account the group self that seemed most relevant to the situation.

Just as people can conceive of any individual as a member of different groups, each group can be described by referring to a variety of traits and behaviors. Nevertheless, people tend to use a limited number of dimensions to specify how their group differs from other groups (*12*). They focus on the ability of their group to achieve desired outcomes (competence), its tendency to approach others in a friendly manner (sociability), and its reliance on important values (morality). When forming an impression of unknown individuals or groups, people first seek out moral information, to determine whether others are honest, reliable, and trustworthy (*13*). The morality of one's group is a primary source of group virtue, even in situations where competence clearly matters (such as when university graduates are considered by potential employers) (*14*).

The group self is activated to make sense of complex social situations. Which of different group selves becomes relevant depends on which is most useful to distinguish the self from others in that context. Whether the group self reflects positively or negatively on the individual depends on how the group compares to other groups, primarily in terms of its perceived morality.

What It Means to Be a Group Member

When the group self is activated, this connects the group's actions and achievements to one's personal experiences and emotions (Fig. 2, row b) (*15*, *16*). As a result, people may experience guilt and shame when confronted with historical wrongdoings of their group, such as harm done to indigenous populations during times of colonization. These emotions are felt more intensely when people feel more self-involved in the group, even if these are triggered by events that happened before they were born (*17*). Social identification processes also make people more self-involved in the emotional experiences of other members of their group. Greater levels of arousal, empathy, and attention are elicited when observing the pleasure and displeasure of those who belong to the same group as the self, as compared to the experiences of members of other groups (*18*). People are more motivated to interpret facial expressions of emotions when displayed by members of their own group than those of other groups and are able to interpret these emotions more accurately (*19*).

The self-reported intensity of emotions may be unreliable or subject to strategic self-presentation. The introduction of physiological and brain imaging techniques in this area of research offers more direct evidence of self-involvement in the group at very early stages of information processing. For instance, people displayed similar patterns of brain activity indicating the experience of emotion when observing members of their group displaying sad facial expressions, as when considering a personal experience that made them feel sad (*20*). They also showed similar motor cortex activity during their own actions as when observing another member of their group performing this action (*21*). Such vicarious activation of neural networks was not found when observing members of another group.

Importantly, evidence of differential brain activity indicating that people can literally think of others in the group as they think about themselves was also obtained with so-called "minimal" groups. Even when people were randomly assigned to racially mixed teams consisting of individuals who did not know each other, functional magnetic resonance imaging revealed more brain activity when viewing novel own-team (versus other-team) faces. The increase in brain activity, which occurred automatically, led people to indicate more liking for individuals who were assigned to their own team, rather than the other team, regardless of their race (*22*).

The activation of the group self also has important consequences for the way in which people perceive and respond to human suffering. The medial prefrontal cortex (mPFC) is a specific brain area associated with the processing of information about other humans. When African Americans and Caucasian Americans observed victims of a natural disaster, they displayed more activity in this brain area when watching members of their own ethnic group. This also predicted how they responded to humans in need: The increased brain activity was related to more empathy and greater willingness

	Review section	Processes	Focus	Implication
a	Making sense of the situation	Social categorization, social comparison	Group salience and distinctiveness	Context dependence of group self
b	What it means to be a group member	Social identification In/outgroup differentiation	Emotional significance of group membership	Inclusion of others in the (group) self
c	Being included or excluded	Categorization threat Stereotype threat	Convergence of own vs. others' definition of self	Transfer of group characteristics to self
d	When the group self is under threat	Individual mobility "Passing"	Implications of stigmatized identity	Self-denial, self-hate
e	Standing up for the group self	Social creativity Social change	Group's standing, collective self-esteem	Emancipation, group commitment

Fig. 2. Understanding the group self: roadmap to the relevant processes, their foci, and their implications, as described in the different sections of this Review.

to donate money and time to help this person (*23*). Different parts of the mPFC are activated depending on how one's own group relates to the group that is represented and the specific emotions this elicits (Fig. 3). Viewing representatives of a group people tend to distance from the self (such as addicts or the homeless) does not even activate this brain area, suggesting that the group self may make others outside the group seem less than human (*24*).

People's self-involvement in groups makes them respond differently to those who belong to their own group than to those who belong to other groups. Studies with so-called minimal groups demonstrate that these effects cannot be explained by interpersonal (e.g., ethnic or gender) similarity, liking, or outcome dependence. Studies of brain activity show that the group self operates automatically and influences the way people perceive and interact with others in ways that cannot be explained by empathic ability or general concern for human suffering.

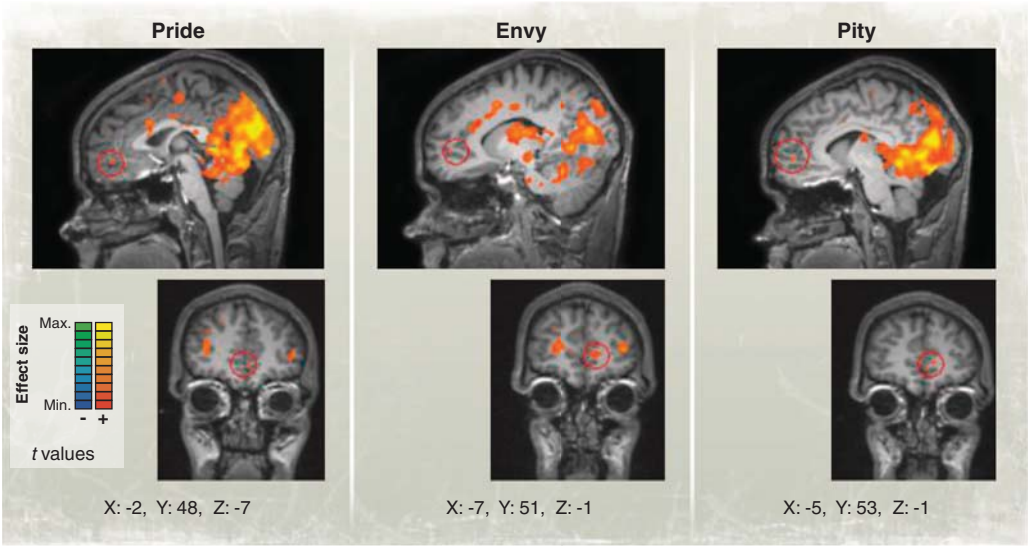


Fig. 3. Results of a subtractive analysis of blood oxygen level–dependent (BOLD) activations while participants viewed pictures of social groups representing the four quadrants of the stereotype content model (SCM). Results for the three SCM cells (pride, envy, and pity) showing significant activation in the mPFC are shown. [Reprinted with permission from figure 1 of (*24*)]

Being Included or Excluded

The group self not only determines how people respond to others but also has an impact on how they prefer others to respond to them (Fig. 2, row c). The desire to be valued by others and included in groups is a basic human need (*25*). However, this does not imply that people relinquish their feelings of individual uniqueness (*26*). Because people all belong to many different groups, they have some leeway to actively select and promote different group selves in different contexts. Unfortunately, the group self that is most easily visible and clearly salient for others (such as ethnicity or gender) is not necessarily what seems most self-relevant at all times. For instance, people may prefer not to be considered in terms of their ethnicity or gender at school or at work, even when they acknowledge that they belong to the group, value the group, or realize it can yield important benefits (*27*).

When people are perceived and treated by others in terms of a group that does not converge with their own preferred group self, this causes feelings of (categorization) threat. When others make unsolicited and (subjectively) inappropriate references to the group self, this can have adverse effects on motivation and achievement (stereotype threat), especially when members of one's group typically do not perform very well in that situation (*28*). The anxiety and distraction raised in this way increase blood pressure (*29*) and engage cognitive resources that are needed for demanding tasks (*30*). As a result, when female students were required to indicate their gender before taking their AP Calculus exam (rather than afterward), their test performance was reduced by 33% (*31*). Similar mechanisms have been found to result in the underperformance of African Americans on GRE tests. Over time, cycles of stereotypic expectations and underperformance can cause a loss of interest in school performance and lower career aspirations.

Invoking the group self in this way can impair the performance of women in leadership, of elderly adults on memory tests, of students with a mental illness on tests of rational thinking, or of whites on measures of racism (*32*). Yet individuals often fail to recognize that they personally suffer from their inclusion in a disadvantaged group (*33*). Indeed, adverse performance effects mainly emerge when references to the group self remain implicit and subtle (as when being asked to indicate one's ethnicity on a test form or when being the only woman in a group of men). When people are more explicitly addressed in terms of a group self they consider irrelevant to the situation, they are better able to resist confirming group-based expectations (*34*).

A lack of convergence between one's preferred definition of self (e.g., as a professional at work) and they way one is treated by others (e.g., as an ethnic minority member) can be psychologically damaging. Longstanding or prototypical group members can be wary of accepting and including those who have a different background (such as second-generation migrants or lower-income students at an Ivy League institution) or individuals who are critical of the group's set ways (such as whistleblowers). This type of social rejection generates brain activity resembling that seen during physical sensations of pain (*35*).

Hazing procedures or (informal) initiation rites test the resolve of new group members and keep them "on their toes" while they learn the ways of the group. When their demonstrations of loyalty and commitment are not rewarded, this makes people feel disrespected and reluctant to continue their cooperation (*36*). Instead, they are easily tempted to display hostility and aggression (*37*). Analyses of political radicalization or school shootings consider these as extreme consequences of the individual's desire to get back at the group that excluded them (*38*). When individuals who do not easily adapt are marginalized or excluded (*39*), the group loses out too. A group that punishes deviance and is not open to dissent has less potential for innovation and change (*40*).

When the Group Self Is Under Threat

Most Western societies tend to consider people as separate individuals who can pursue their own goals, depending on their competences and achievements. This creates an "illusion of meritocracy," suggesting that those who belong to disadvantaged or devalued groups should just break away from the group to improve their own outcomes (Fig. 2, row d) (*41*). This is easier said than done, if only because others will not necessarily accept or acknowledge those who try to pass into another group, as indicated above.

When group membership is immediately visible, emphasizing differences between the self and other group members may seem the best

way to escape negative implications of the group self. Individuals who follow this strategy often suffer, because those who violate stereotypic expectations tend to raise scorn and hostility. When Hillary Clinton controlled her emotions, she was called an "ice queen." Condoleezza Rice was accused of being an "Oreo cookie" when pursuing her political ambitions: black on the outside, but white inside. The group may also suffer when others in the group are put down in an attempt to lift up the self. For instance, when senior policewomen tried to escape sexist treatment at work by emphasizing their masculine qualities, they also denied that other women in the organization suffered from discrimination (*42*).

When one's group membership is less immediately visible, people may try to avoid revealing that they are part of a devalued group, as a way to improve their personal outcomes. Those with a history of mental or physical illness can avoid talking about this period in their lives or make up excuses to hide medical examinations or treatments. Some homosexuals bring cross-sex dates to social events to escape questions about their sexual orientation. Many prefer not to reveal their religious or political preferences at work. In his novel *The Human Stain*, Philip Roth shows the extreme implications this can have, when his character Coleman Silk breaks all ties with his family to avoid revealing even to his wife and children that he is of African-American descent.

People do this because it seems an efficient way to escape the disadvantages of group-based stigma. But living a lie is not without cost to emotional and physical well-being. Self-denial can turn into self-hate. Failing to reveal who they really are or where they come from deprives people of the support and esteem of others like them (*43*). The increased mindfulness they invest to monitor what is being said to whom depletes cognitive resources needed for intellectual tasks (*44*). Homosexuals who fear being exposed may fail to acquire the health care they need (*45*).

Standing Up for the Group Self

When the group self is under threat, people may also decide to stand up for the group (Fig. 2, row e). They will fight to improve their group's outcomes or image when sufficiently committed to the group. This may be the case either because they care for the group or because they realize they cannot easily escape being associated with the group (*46*). Different strategies are available to harness the group self. People adapt their efforts to the situation and its likely opportunities for change.

A lack of resources and political power can make it unlikely that an actual improvement in the societal position of the group can be realized in the near future. If achieving social change is not feasible, people may still redefine the group self in a positive way to increase their collective self-esteem. They can engage in so-called social creativity by emphasizing characteristic achievements of their group, for instance, in music or sports; comparing their group with another group that is even worse off; or reevaluating existing group features ("black is beautiful"; "gay pride"). When status relations between groups are less secure, as is the case with sports teams or political factions, groups will be tempted to compete with each other to demonstrate their superior skill and achievements or to gain the upper hand in obtaining public support and access to resources as a way to improve their group's standing.

Research examining when people stand up for the group self has mainly focused on emancipation attempts by members of disadvantaged groups. Nevertheless, their attempts at social creativity or social change may in turn threaten and mobilize others to protect their own group's existing privileges. Attempts to empirically assess these dynamics of relations between social groups have suffered from the ambiguity of self-reported experiences and intentions. Even when explicit statements reflect the desire to appear unprejudiced toward members of other groups, implicit indicators may reveal nonverbal displays of hostility toward them (*47*). In a similar vein, acknowledging that one's group is not very successful may either reflect a realistic awareness of current outcomes or indicate a lack of confidence in the group's ability to improve.

Recent research has started to use cardiovascular reactivity as a more dynamic indicator of the way people respond to and cope with the plight of their group. Specific patterns of change in heart rate and blood pressure can be used to distinguish between the emergence of positive versus negative arousal in members of advantaged and disadvantaged groups. Negative arousal indicates a state of threat, whereas positive arousal indicates a state of challenge and confidence in the ability to adequately cope with the situation. These autonomic responses reveal that belonging to a group with low status is threatening as long as differences between groups seem stable. However, the group's unfavorable position becomes a source of positive challenge when there is a possibility for change. Conversely, members of groups with high status experience threat when they realize that their group's current status is unstable (*48*).

Cardiovascular indicators also help ascertain whether people respond differently to taxing situations, depending on whether they rely on the group self or the individual self. This was demonstrated in a study where women performed a computer task in which they were required to parallel park a car into a tight spot. Those who focused on the individual self (low–gender-identified women) were positively challenged when their individual self-worth had been affirmed. Those who focused on the group self (high–gender-identified women) experienced challenge only after having been reassured of their group's worth (*49*).

For members of advantaged groups, changing status relations raise insecurity, even when they are convinced that such change would be legitimate. Individuals who indicate liking and positive attitudes toward a devalued group display cardiovascular patterns of threat when asked to cooperate with members of that group and perform suboptimally during this cooperation (*50*). After interacting with a black experimenter, white participants displayed more brain activity indicating cognitive control when responding to black and white faces. However, these efforts to suppress automatic bias depleted their cognitive resources and impaired their subsequent task performance (*51*).

Conclusions and Implications

Efforts at documenting universal principles of human behavior have tended to focus on the individual as the center of self-awareness, self-control, and self-concerns. Such personal concerns or idiosyncratic preferences may seem less important when people feel part of a group. The group self can make people engage in behavior that opposes personal ideals (do no harm) when this seems the only way to acquire important group goals (so that violence seems justified to achieve social change). Likewise, personal interests (such as care for one's own safety) can be set aside in the pursuit of shared ideals (patriotism). Acknowledging this helps advance scientific inquiry and benefits those who work with groups in practice.

Those who manage others at work may already know that loyalty and commitment to team and organizational goals can be increased when individual workers feel emotionally connected to each other. But they should also realize that the practice of rewarding employees for their personal achievements makes it more difficult to forge such connections, as it fosters competition between co-workers instead of appealing to the group membership that they share.

Members of different groups in society too may benefit by learning about the group self. Our reliance on moral judgments to convey the worth of different groups makes it tempting to consider our own group and its characteristic values as morally superior. Groups that endorse different moral values can elicit hostility (*52*). Their reluctance to share our moral values makes them seem less human, which justifies aggression against them (*53*). Understanding the implications of differences in moral values for the group self offers a new perspective on such conflicts and on interventions that will be most effective to resolve them.

A broader group-level approach to socio-economic differences in the world can also inform social policies aiming to relieve human

suffering. Attempts to raise empathy and support tend to provide information about the plight and unfair outcomes of others. However, clarifying how people are connected to those in need and appealing to the group self may be more effective than relying on individual displays of altruism.

This is not to say that it is easy to appeal to a particular group self. Because of the complexity of contemporary multicultural societies, there is often considerable ambiguity as to which group self is at stake. Yet there is added value in trying to understand which group self people find self-relevant. People suffer when others approach them as representatives of a group they do not see as self-defining. The resulting loss of control over their own sense of self can severely affect psychological well-being, academic and professional success, and physical health. Marginalization and disillusionment can be prevented when one's preferred ways of self-definition are respected by others.

Being acknowledged and valued for what they are gives people the confidence to fit in and meet new requirements. Excluding those who are different from the majority or critical of its ways is costly. At best, this prevents individuals from achieving their full potential. At worst, this can result in extremism and social instability, because they no longer care about a society that does not accept them.

Acknowledging that people can have multiple group selves can help reduce tensions between one's own and others' conceptions of self. Respecting and valuing the group self that they find important provides people with the confidence to learn and the motivation to embrace other group selves too (*54*).

Even when consensual beliefs encourage and expect individuals to take a position in society that reflects their personal merit, achieving this is not always easy. Strategies that can be used to escape negative expectations associated with the group of origin can come at great cost for the individual as well as for other members of the group. More often than not, the achievement of individual potential is possible only when members of disadvantaged groups successfully manage their devalued group identity. This requires more than simply learning appropriate skills, acquiring proper manners, or adapting one's speech and dress style. Changing who you are and where you belong can easily be interpreted as a sign of rejection by those who stay behind. This puts ambitious individuals from a disadvantaged background in a tenuous position and makes it clear that they need explicit encouragement and support from others to be able to succeed.

Changing relations between groups in society are a source of stress and anxiety. Members of disadvantaged groups can be insecure about their ability to achieve real change. Members of advantaged groups can be insecure about whether they are able to suppress and overcome their biased expectations. Even with the best of intentions, this is emotionally and cognitively taxing. Bringing together individuals from different groups does not necessarily help overcome these concerns. As long as the group self is activated, such interactions may paradoxically exacerbate mutual feelings of discomfort, which make it difficult to get along. In cases as these, searching for commonalities that may forge a shared group identity offers a more productive way forward (*55*).

As much as the group self can be a source of support and inspiration, it can also undermine self-confidence and raise societal tensions. Reminding people of their personal goals and individual responsibilities is not the most effective way to manage these liabilities. Bringing out the best in each individual requires that we acknowledge and celebrate the group self.

References and Notes

1. A demonstration of how people start behaving differently when placed in a particular group can be found at www.bbcprisonstudy.org/index.php.
2. J. Jetten, C. Haslam, S. A. Haslam, Eds., *The Social Cure: Identity, Health and Well-Being* (Psychology Press, New York, 2012).
3. M. R. Leary, J. P. Tangney, *Handbook of Self and Identity* (Guilford Press, New York, 2003).
4. H. Tajfel, J. C. Turner, in *The Social Psychology of Intergroup Relations*, W. G. Austin, S. Worchel, Eds. (Brooks/Cole, Monterey, CA, 1979), pp. 33–47.
5. T. Postmes, N. R. Branscombe, Eds., *Rediscovering Social Identity: Core Sources* (Psychology Press, New York, 2010), pp. 341–356.
6. S. A. Haslam, S. D. Reicher, M. J. Platow, *The New Psychology of Leadership: Identity, Influence, and Power* (Psychology Press, New York, 2011).
7. N. Ellemers, S. A. Haslam, in *Handbook of Theories of Social Psychology*, P. van Lange, A. Kruglanski, T. Higgins, Eds. (Sage, London, 2011), pp. 379–398.
8. C. W. Leach *et al.*, *J. Pers. Soc. Psychol.* **95**, 144 (2008).
9. Examples of tests that are used to examine the implicit associations that emerge when thinking of particular groups can be found at https://implicit.harvard.edu/implicit/demo.
10. J. C. Turner, K. J. Reynolds, in *Handbook of Theories of Social Psychology*, P. van Lange, A. Kruglanski, T. Higgins, Eds. (Sage, London, 2011), pp. 399–417.
11. D. M. Grace, B. J. David, M. K. Ryan, *Child Dev.* **79**, 1928 (2008).
12. S. T. Fiske, A. J. C. Cuddy, P. Glick, J. Xu, *J. Pers. Soc. Psychol.* **82**, 878 (2002).
13. M. Brambilla, S. Sacchi, P. Rusconi, P. Cherubini, V. Y. Yzerbyt, *Br. J. Soc. Psychol.* **51**, 149 (2012).
14. C. W. Leach, N. Ellemers, M. Barreto, *J. Pers. Soc. Psychol.* **93**, 234 (2007).
15. E. R. Smith, C. R. Seger, D. M. Mackie, *J. Pers. Soc. Psychol.* **93**, 431 (2007).
16. V. Y. Yzerbyt, in *Bridging Social Psychology. Benefits of Transdisciplinary Approaches*, P. A. M. van Lange, Ed. (Lawrence Erlbaum Associates, Mahwah, NJ, 2006), pp. 391–396.
17. N. R. Branscombe, B. Doosje, Eds. *Collective Guilt* (Cambridge Univ. Press, Cambridge, 2004).
18. L. M. Brown, M. M. Bradley, P. J. Lang, *Biol. Psychol.* **71**, 303 (2006).
19. S. G. Young, K. Hugenberg, *J. Pers. Soc. Psychol.* **99**, 964 (2010).
20. J. N. Gutsell, M. Inzlicht, *Soc. Cogn. Affect. Neurosci.* (2011).
21. J. N. Gutsell, M. Inzlicht, *J. Exp. Soc. Psychol.* **46**, 841 (2010).
22. J. J. Van Bavel, D. J. Packer, W. A. Cunningham, *Psychol. Sci.* **19**, 1131 (2008).
23. V. A. Mathur, T. Harada, T. Lipke, J. Y. Chiao, *Neuroimage* **51**, 1468 (2010).
24. L. T. Harris, S. T. Fiske, *Psychol. Sci.* **17**, 847 (2006).
25. R. F. Baumeister, M. R. Leary, *Psychol. Bull.* **117**, 497 (1995).
26. M. B. Brewer, *Pers. Soc. Psychol. Bull.* **17**, 475 (1991).
27. N. Ellemers, R. Spears, B. Doosje, *Annu. Rev. Psychol.* **53**, 161 (2002).
28. C. M. Steele, *Am. Psychol.* **52**, 613 (1997).
29. J. Blascovich, S. J. Spencer, D. Quinn, C. Steele, *Psychol. Sci.* **12**, 225 (2001).
30. T. Schmader, M. Johns, C. Forbes, *Psychol. Rev.* **115**, 336 (2008).
31. K. Danaher, C. S. Crandall, *J. Appl. Soc. Psychol.* **38**, 1639 (2008).
32. J. R. Shapiro, A. M. Williams, *Sex Roles* **66**, 175 (2012).
33. K. A. Quinn, N. J. Roese, G. L. Pennington, J. M. Olson, *Pers. Soc. Psychol. Bull.* **25**, 1430 (1999).
34. L. J. Kray, L. Thompson, A. Galinsky, *J. Pers. Soc. Psychol.* **80**, 942 (2001).
35. N. I. Eisenberger, M. D. Lieberman, K. D. Williams, *Science* **302**, 290 (2003).
36. T. R. Tyler, *Res. Organ. Behav.* **21**, 201 (1999).
37. K. D. Williams, *Annu. Rev. Psychol.* **58**, 425 (2007).
38. M. R. Leary, R. M. Kowalski, L. Smith, S. Phillips, *Aggress. Behav.* **29**, 202 (2003).
39. J. M. Marques, D. Paez, *Eur. Rev. Soc. Psychol.* **5**, 37 (1994).
40. J. Jetten, M. J. Hornsey, Eds., *Rebels in Groups: Dissent, Deviance, Difference, and Defiance* (Blackwell, Oxford, 2011).
41. N. Ellemers, C. Van Laar, in *Handbook of Prejudice, Stereotyping, and Discrimination*, J. F. Dovidio, M. Hewstone, P. Glick, V. Esses, Eds. (Sage, London, 2010), pp. 561–576.
42. B. Derks, C. Van Laar, N. Ellemers, K. de Groot, *Psychol. Sci.* **22**, 1243 (2011).
43. D. E. S. Frable, L. Platt, S. Hoey, *J. Pers. Soc. Psychol.* **74**, 909 (1998).
44. L. Smart, D. M. Wegner, in *The Social Psychology of Stigma*, T. F. Heatherton, R. E. Kleck, M. R. Hebl, J. G. Hull, Eds. (Guilford Press, New York, 2000), pp. 220–242.
45. S. D. Young, B. Monin, D. Owens, *Health Psychol.* **28**, 675 (2009).
46. N. R. Branscombe, M. T. Schmitt, R. D. Harvey, *J. Pers. Soc. Psychol.* **77**, 135 (1999).
47. J. F. Dovidio, K. Kawakami, S. L. Gaertner, *J. Pers. Soc. Psychol.* **82**, 62 (2002).
48. D. Scheepers, *J. Exp. Soc. Psychol.* **45**, 228 (2009).
49. B. Derks, D. Scheepers, C. Van Laar, N. Ellemers, *J. Exp. Soc. Psychol.* **47**, 178 (2011).
50. W. B. Mendes, J. Blascovich, B. Lickel, S. Hunter, *Pers. Soc. Psychol. Bull.* **28**, 939 (2002).
51. J. A. Richeson *et al.*, *Nat. Neurosci.* **6**, 1323 (2003).
52. E. Mullen, L. J. Skitka, *J. Pers. Soc. Psychol.* **90**, 629 (2006).
53. S. Opotow, *J. Soc. Issues* **49**, 71 (1993).
54. B. Derks, C. Van Laar, N. Ellemers, *Social Issues Policy Rev.* **1**, 217 (2007).
55. E. Bijleveld, D. Scheepers, N. Ellemers, *PLoS ONE* **7**, e33681 (2012).

Acknowledgments: This work was supported by the Leiden University Institute of Psychological Research. I thank D. de Gilder and two anonymous reviewers for their helpful comments.

10.1126/science.1220987

PERSPECTIVE

ADAPTING TO A MULTICULTURAL FUTURE

Richard J. Crisp* and Rose Meleady

Humans have an evolved propensity to think categorically about social groups. This propensity is manifest in cognitive processes that have broad implications for public and political endorsement of multicultural policy. Drawing on these principles, we postulate a cognitive-evolutionary account of human adaptation to social diversity. This account explains broad social trends marking a resistance to multiculturalism, while providing an important reorienting call for scholars and policy-makers seeking intervention-based solutions to the problem of prejudice.

Our lifetimes have witnessed unprecedented levels of intercultural exchange, enabled by ever-enhancing transport technologies and a truly globalized communications infrastructure. We no longer live in the homogenized, provincial worlds that have typified much of human history. Nationality no longer denotes race as Western societies are increasingly characterized by complex combinations of national, religious, ethnic, and cultural identity (*1*). In terms of our evolutionary history, we are entering uncharted territory: Never before have humans had to deal with social environments defined by such degrees of diversity and differentiation.

Unsurprisingly, how people adapt to this diversity has been a persistent and pervasive characteristic of public, political, and scholarly debate. Proponents of the "multiculturalism hypothesis" assert that societies valuing a constellation of distinct cultural, ethnic, and religious identities inspire intergroup harmony (*2*)—an assertion well founded on empirical research (*1*, *2*). Yet such findings stand in stark contrast to the realpolitik of intercultural relations. Despite state-endorsed policies promoting multiculturalism, diversity often appears to accompany conflict, not peaceful coexistence. A striking example is the multiracial city of Bradford, (Yorkshire, UK), which in July 2001, despite high levels of geographical integration of the south Asian and white communities, was the epicenter of Britain's worst-ever race riots (*3*). It is therefore perhaps not surprising that opponents of multiculturalism argue that diversity is the anathema to social cohesion and is integral to the etiology of civic unrest and social disharmony.

Recent reviews of the literature have gone on to question the assertion that closer social integration promotes positive outcomes (*4*). Experimental studies add to this complex picture, showing that attempts to promote multicultural ideology can reduce prejudice while at the same time maintaining (often negative) cultural stereotypes (*5*) or buttressing support for inequality (*6*).

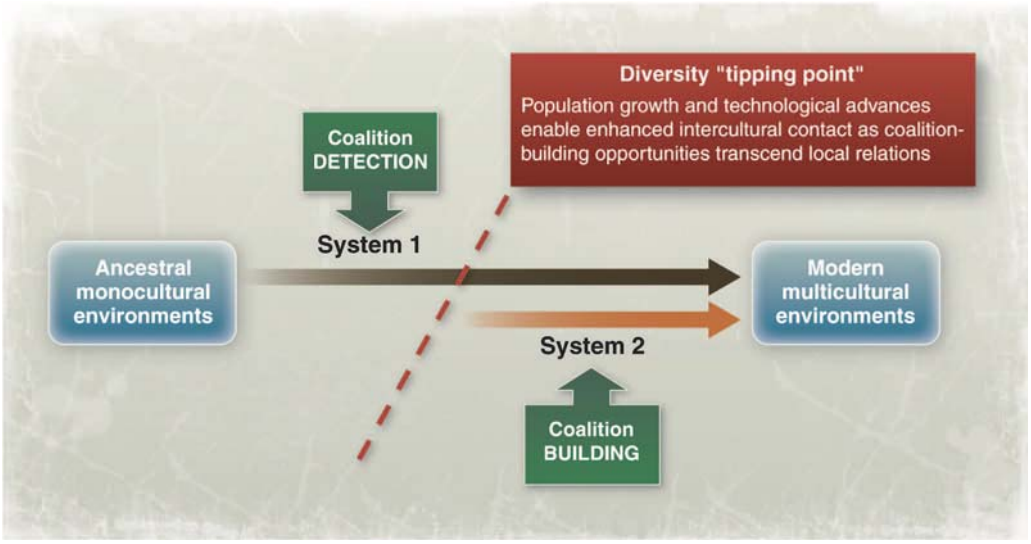


Fig. 1. An evolved dual-system computational architecture for detecting and building coalitions. System 1 is an early-evolved system for automatically detecting covariance with category cues, adaptive in largely monocultural ancestral environments. Coalitional adaptation to multicultural environments requires System 2 to resolve category-based inconsistencies arising from the complex and cross-cutting affiliations characteristic of social diversity.

Behavioral scientists (ourselves included) often bemoan the glacial rate at which basic social science is translated into policy and practice. Ironically, it is the perplexing picture painted by existing diversity perspectives that is in danger of defining the field's policy impact. In the absence of any clear message from the scientific community, policies that espoused the multicultural ideal over the past 50 years have begun to unravel, with politicians and policy-makers declaring the multicultural "experiment" a failure. As sociologist Joppke (*7*) observed, after many years characterized by the liberal pursuit of diversity-based ideals, we are now witnessing a chronic lack of public and political support that has precipitated a wholesale retreat of the multicultural state. Joppke's observations resonate in political rhetoric. Echoing comments made by the British prime minister (*8*), German Chancellor Angela Merkel recently announced that attempts to build a multicultural society in Germany have "utterly failed" and that the so-called "multikulti" concept, where people would "live side-by-side" happily, simply did not work (*9*).

Are we therefore, as some scholars suggest, heading irrevocably toward a "clash of civilizations" (*10*), as tensions erupt and escalate around differences in race, religion, culture, and ethnicity? We contend that this is not the case and that such a conclusion ignores a raft of recent studies in the behavioral sciences, and evolutionary social cognition in particular, that provide a more panoramic perspective to the multicultural debate. Most important, consideration of this work changes the fundamental question asked of multicultural ideals and suggests a need to reorient scholarly, public, and political foci on this most critical of contemporary social issues.

To fully understand ailing support for multicultural policy, we must ask why people are resistant to social diversity in the first place. The answer may lie in evolved human preferences for homogeneity, stability, simplicity, and structure (*11*). These preferences are exemplified in mental systems for representing groups and group differences. Adaptive fitness is contingent on the development of

Centre for the Study of Group Processes, School of Psychology, University of Kent, Canterbury, Kent CT2 7NP, UK.

*To whom correspondence should be addressed. E-mail: r.crisp@kent.ac.uk

functionalized cognitive architectures that most effectively represent the prevailing social ecology (*12*). Accordingly, groups should provide the context for the selection and evolution of cognitive processes—the natural way for the mind to construe reality for an inherently social species (*11*).

Consistent with this explanatory framework, human social cognition has adapted to ancestral environments defined by simple "us" versus "them" category boundaries, enabling a clear way of distinguishing friend from foe. This preference for representing categorical differences is reflected in contemporary social psychological research. Extensive evidence shows how categorical thinking enables others to be identified, in a matter of milliseconds, as ingroup or outgroup members (*13*). These findings are indicative of an evolved neurocomputational architecture needed to rapidly detect covariation between allegiance and perceptual categorization cues (*14*).

This sensitivity to group differences was an adaptive, efficient coalitional detection system in largely localized and monocultural ancestral environments. In such environments, tribal boundaries were likely to be clear-cut, with little potential for category confusion or need to construct cross-cutting tribal allegiances. It is a mechanism, however, rather unsuited to the level of diversity that, in a short evolutionary time scale, has come to characterize many modern societies. Processing preferences for simple "us" versus "them" category boundaries are antithetical to environments in which nationality, race, religion, and ethnicity form multiple, cross-cutting bases for social affiliation. It is the conflict between this changing social ecology and evolved preferences for defined category boundaries that can explain generalized resistance to multicultural ideologies, avoidance of intercultural contact, and negative reactions to social diversity.

However, heuristic systems must possess the ability to make modifications to extant cognitive structures. This is essential to improve efficient classification in future encounters—that is, to deal with exceptions to the rule. The operation of this updating mechanism may be the key to understanding how humans successfully adapt to social diversity.

Our contention is that two systems now constitute humans' coalitional machinery (Fig. 1). System 1 is an automatic, early-evolved system for detecting covariance with "us" versus "them" category cues, designed for the effective maintenance of existing allegiances in largely monocultural ancestral environments (*14*). However, with population growth and advances in transport technology, opportunities for intercultural contact began to transcend local intergroup relations. In this new world of multiple intergroup differences (and potential threats), a second system was needed to capitalize on more complex, cross-cutting bases for coalitional security. Such a system provides an adaptive advantage—a cognitive capability for building new alliances in a world otherwise overrun by outgroups.

The processing steps involved in this second system are specified in a recent model of diversity-based cognition (*15*): (i) Category-based inconsistency is detected (i.e., positive and/or counterstereotypical outgroup behaviors), which (ii) initiates inhibition of existing category representations in favor of more generative thinking, which (iii) results in a creative reconstrual of targets (as potential coalition allies). This cognitive resolution of category-discrepant targets enabled ancestral humans to reach beyond existing intergroup differences—to turn these intergroup differences into intragroup diversity, competition into coalition. The complexity of the computational processes involved in these steps is considerable but is consistent with what we know about the link between brain size and the evolution of human societies (*12*). Indeed, it can be argued that the human brain is the largest among primates precisely because it is specialized to deal with the challenging forms of social interaction involved in intercultural contact.

Thus, while humans are evolutionarily disposed to think heuristically about category boundaries (System 1), they also possess the computational mechanics that allow a bypassing of this system when it is necessary to update and revise these representations (System 2). Diversity-based applications of this second system provide an explanatory framework for macrosocial trends rooted in the social cognition of the individual. Consistent with the processing principles outlined above, successful intercultural contact requires cognitive control to counter evolved tendencies to contrast ingroups from outgroups. At the most basic level, positive outgroup encounters will be inconsistent with existing category representations (outgroup = negative), so the second system is triggered to inhibit, resolve, and reclassify.

Evidence for the operation of this second system can be found in experimental studies of social categorization (*16*, *17*). For instance, these studies have shown that when outgroup targets simultaneously share a basis for mutual affiliation with the ingroup (i.e., conflicting categorization cues), this leads to more positive perceptions. However, even targets defined by multiple (e.g., five) outgroup categories are more positively evaluated than targets defined by a single outgroup identity—a tendency statistically explained by a cognitive reconstrual of the target as an individual rather than an (oppositional) group member. This is particularly the case when multiple outgroup categories are low in conceptual overlap (e.g., German, Muslim) versus highly correlated (e.g., Pakistani, Muslim). Such findings support the idea that it is the inconsistency between multiple categorizations that initiates System 2. These reconstruals of outgroup targets consume attentional reserves (*18*), but this is an adaptation that discourages overengagement. In ancestral environments, it would be unwise to too easily cast aside category boundaries learned through experience.

In sum, experimental research supports the proposition that coalitional architectures comprise two systems: one for automatically detecting covariance with category cues (*14*), and one for updating expectancies in environments characterized by social diversity (*15*). In multicultural environments, the second system is required to inhibit existing category-based expectations and compute potential risks, costs, and benefits arising from cooperative interaction. Such is the way complex societies are formed, fostered, prosper, and grow.

This cognitive-evolutionary account has implications for both future research and social policy initiatives. The existence of two coalitional systems could explain diverging outcomes after intercultural contact, as well as withdrawal from ideologies promoting greater integration. Simply bringing groups together will likely do little other than initiate System 1, a mechanism for maintaining existing coalitional boundaries. In contrast, when policy encourages individuals to embrace new, cross-cutting bases for social affiliation, particularly those that defy category-based expectancies, then activation of System 2's coalition-building function may more readily reduce the negative emotions that inhibit intergroup contact, and in turn promote more positive engagement with outgroups.

Understanding when either system will be initiated could also help in the development of diversity-based interventions for reducing prejudice. For instance, the model suggests that future research should not simply aim to discourage the use of System 1 (e.g., through colorblind approaches) but to actively engage System 2 (e.g., recognize cross-cutting or counterstereotypic combinations of gender, race, occupation, and socioeconomic status). This idea—that coalitional systems can adjust to include outgroups—affirms and extends existing social cognitive theory advocating the recognition of group differences (*19*).

Incorporating the principle of cognitive adaptation also provides the potential to leverage new paradigms from evolutionary psychology in the pursuit of prejudice reduction. For instance, evolutionary theory regarding the hierarchical nature of human motivation (*20*) suggests that to facilitate activation of System 2, interventions should be framed by the goal of coalition building while avoiding other adaptive goals such as (in)group protection.

This cognitive-evolutionary account also provides the basis for greater cross-disciplinary integration in the development of diversity theory, involving contributions from both the biological and social sciences. For instance, local econom-

ic and political dynamics, immigration history, and the pace of intercultural mixing should interact to predict activation of brain regions linked to the operation of System 2, which should correspondingly be reflected in macrosocial trends toward tolerance and harmonious intergroup relations.

Has the multicultural "experiment" failed? The answer is that we just don't know, not because science has failed to make a contribution, but because we have not been asking quite the right questions. The research discussed above suggests that social diversity results in coalitional (i.e., cooperative, tolerant) behavior only when it requires one to think beyond the ancestral building blocks of modern society. Predicting how people will react to multiculturalism therefore requires more focused scientific research on which coalitional systems are activated when different policies are communicated (or communicated in different ways).

More generally, we would argue that to provide clear science-based directives concerning multicultural policy, researchers must better understand the operation of cognitive systems that manage our capacity to navigate socially diverse environments. We must harness those systems and capitalize on their potential to inform, focus, and frame future policy in this area. Perhaps only by considering the origins and evolution of such systems can we fully understand the impact of broad-based initiatives for social change.

References

1. V. C. Plaut, *Psychol. Inq.* **21**, 77 (2010).
2. M. Verkuyten, *J. Pers. Soc. Psychol.* **88**, 121 (2005).
3. T. Cantle, *Community Cohesion: A Report of the Independent Review Team* (Home Office, London, 2001).
4. F. W. Rudmin, *Rev. Gen. Psychol.* **7**, 3 (2003).
5. C. Wolsko, B. Park, C. M. Judd, B. Wittenbrink, *J. Pers. Soc. Psychol.* **78**, 635 (2000).
6. K. R. Morrison, V. C. Plaut, O. Ybarra, *Pers. Soc. Psychol. Bull.* **36**, 1648 (2010).
7. C. Joppke, *Br. J. Sociol.* **55**, 237 (2004).
8. www.bbc.co.uk/news/uk-politics-12371994
9. www.bbc.co.uk/news/world-europe-11559451
10. S. P. Huntington, *Foreign Aff.* **72**, 22 (1993).
11. L. R. Caporael, *Pers. Soc. Psychol. Rev.* **1**, 276 (1997).
12. R. I. Dunbar, S. Shultz, *Science* **317**, 1344 (2007).
13. C. Stangor, L. Lynch, C. Duan, B. Glas, *J. Pers. Soc. Psychol.* **62**, 207 (1992).
14. L. Cosmides, J. Tooby, R. Kurzban, *Trends Cogn. Sci.* **7**, 173 (2003).
15. R. J. Crisp, R. N. Turner, *Psychol. Bull.* **137**, 242 (2011).
16. R. J. Crisp, M. Hewstone, M. Rubin, *Pers. Soc. Psychol. Bull.* **27**, 76 (2001).
17. N. R. Hall, R. J. Crisp, *Pers. Soc. Psychol. Bull.* **31**, 1435 (2005).
18. R. R. C. Hutter, R. J. Crisp, G. W. Humphreys, G. M. Waters, G. Moffitt, *Group Process. Intergroup Relat.* **12**, 673 (2009).
19. B. Park, C. M. Judd, *Pers. Soc. Psychol. Rev.* **9**, 108 (2005).
20. D. T. Kenrick, S. L. Neuberg, V. Griskevicius, D. V. Becker, M. Schaller, *Curr. Dir. Psychol. Sci.* **19**, 63 (2010).

10.1126/science.1219009

PERSPECTIVE

RELIGIOUS AND SACRED IMPERATIVES IN HUMAN CONFLICT

Scott Atran[1,2,3]* and Jeremy Ginges[4]*

Religion, in promoting outlandish beliefs and costly rituals, increases ingroup trust but also may increase mistrust and conflict with outgroups. Moralizing gods emerged over the last few millennia, enabling large-scale cooperation, and sociopolitical conquest even without war. Whether for cooperation or conflict, sacred values, like devotion to God or a collective cause, signal group identity and operate as moral imperatives that inspire nonrational exertions independent of likely outcomes. In conflict situations, otherwise mundane sociopolitical preferences may become sacred values, acquiring immunity to material incentives. Sacred values sustain intractable conflicts that defy "business-like" negotiation, but also provide surprising opportunities for resolution.

Humans will kill and die not only to protect their own lives or defend kin and kith, but for an idea—the moral conception they form of themselves, of "who we are". Across history, religion has often served to bind members of a primary group, creating moral communities under the rule of immaterial but physically sentient and powerful deities—that is, entities whose semantic description is contradictory and conceptually surprising. Their miraculous features are attention-arresting and memorable, favoring cultural transmission and survival of religious beliefs (*1*). Although logically and empirically inscrutable, such beliefs can be tied to clear norms and behaviors through context-specific interpretations (as with understanding metaphors, and as in weekly sermons). Even when context fixes meaning, seemingly contrary evidence seldom undermines religious belief, especially among groups welded by costly commitment in the face of outside threats (*2*). Belief in gods and miracles also intensifies when people are primed with awareness of death, or when facing danger, as in wartime (*3*). Cross-national analyses show that a country's devotion to a world religion correlates positively with existential insecurity (*4*).

Humans acquire novel beliefs from cultural models (experts, leaders), similarly to how children acquire unfamiliar tastes from adults, by inferring commitment through actions (*5*). Unlike mundane commonsense or scientific beliefs, faith in religious beliefs rests not on logical coherence and empirical evidence but is sustained by costly rituals whose elements may have no active or useful relationships in everyday life (*6*–*8*). Supernatural agents that incentivize costly sacrifices for symbolic beliefs (which reliably identify cooperators and cannot be undermined by reason) favor the cultural linkage, survival, and spread of religious beliefs and rituals.

Studies suggest that costly and seemingly arbitrary ritual commitment to apparently absurd beliefs deepens trust, galvanizing group solidarity for common defense and blinding members to exit strategies (*9*). By contrast, fully reasoned social contracts that regulate individual interests to share costs and benefits of cooperation can be more liable to collapse: Awareness that more advantageous distributions of risks and rewards may be available in the future makes defection more likely. Thus, even ostensibly secular nations and transnational movements usually contain important quasi-religious rituals and beliefs: from sacred songs and ceremonies, to postulations that providence or nature bestow equal rights (*10*). These sacred values act as moral imperatives that inspire nonrational sacrifices in cooperative endeavors and war, generating outsize commitment in low-power groups to resist and often prevail against materially stronger foes (*10*, *11*).

Evolution of Religion

Controversy over religion in human conflict has recently centered on debates about religion's origin and evolution. One view is that religion, including self-sacrifice for sacred group values, is biologically adapted for ingroup cooperation in a competitive environment (*12*). Even arbitrary assignments of sacredness, whose transgressions are taboo and often severely punished, unequivocally distinguish otherwise closely associated groups: Sabbath and Kosher laws isolated Hebrew hill tribes from Canaanite neighbors and,

[1]CNRS–Institut Jean Nicod, 29 rue d'Ulm, 75005 Paris, France. [2]University of Michigan, Ann Arbor, MI 48106, USA. [3]John Jay College of Criminal Justice, New York, NY 10019, USA. [4]New School for Social Research, 80 5th Avenue, 7th Floor, New York, NY 10011, USA.

*To whom correspondence should be addressed. E-mail: satran@umich.edu, gingesj@newschool.edu

in later centuries, diaspora Jews from surrounding societies; ritual association with chosen animals and plants socially separated contiguous Australian aboriginal groups for millennia. The more antagonistic a group's neighborhood, the more proprietary the group's sacred values, increasing ingroup reliance, but also disbelief and potential conflict toward other groups (*7*).

Religious rituals build ingroup solidarity. Experiments in Israeli kibbutzim (cooperatives) show that people who participate most in collective religious ritual more likely cooperate with anyone in their cooperative (*7*). Among 83 utopian communes in the 19th century, religious groups with more costly rituals survived longer than groups with fewer (*7*). Ties between religious ritual and ingroup solidarity may translate into greater willingness to kill and die for a cause in intergroup conflict (*13*). Studies of 60 small-scale societies reveal that groups in highly competitive socioecologies (with frequent warfare) endure the costliest rites (genital mutilation, scarification, etc.) to display solidarity (*7*). Cross-cultural research indicates that participation in collective religious ritual increases parochial altruism and, in relevant contexts, support for suicide attacks (*13*).

An alternative view is that religion (unlike, say, language) involves no particular biological adaptation with clear units of selection. Religion is a fuzzy category with no transparent distinction between beliefs or actions as religious or not (*1*, *8*). Nevertheless, readily identifiable clusters of empirically and logically inscrutable beliefs reoccur cross-culturally as a by-product of nonreligious cognitive functions evolved for mundane purposes. For example, notions of invisible supernaturals may be the by-product of a biologically specialized "mental module" that underlies universal folk psychology, and which is hair-triggered to detect intentional agents as friend or foe under uncertain conditions: People easily imagine voices in the wind or faces in shadows because, from an evolutionary vantage, "better safe than sorry" (*14*).

"New Atheist" thinkers who embrace by-product views consider religion an intellectually worthless but dangerous evolutionary accident readily exploited by predatory societies and kleptocrats (*15*). But most by-product theorists now focus on merging by-product views with adaptationist accounts emphasizing cultural (rather than biological) selection (e.g., by emulating historical success) (*6*, *9*). The thesis is that once belief in supernatural agency emerged as a by-product of mundane cognitive processes, cultural evolution favored the spread of high gods who impose inviolable rules and costly, repetitive rituals upon abstract moral communities, prescribing what people should and shouldn't do (*8*).

As human groups expand into resource-rich environments, they tend toward greater competition, and thrive by becoming larger. But more opportunities arise for larger groups to fissure. Transcendent supernatural agents help solve this problem by providing absolute and context-free moral authority to elicit cooperation among even anonymous individuals (*9*). Priming people with belief in powerful, all-seeing deities (e.g., oaths on a Bible) induces fairness to strangers and self-monitoring (which reduces collective burden in punishing misbehavior) (*6*).

Archaeology reveals clear, coevolutionary connections between religion, ritual, and complex societies. Finds from Mesopotamia and Mesoamerica indicate that rituals became more elaborate and costly as societies developed from foraging bands into chiefdoms and states (*16*). Contrary to notions of religion or god as a direct product of a biologically adapted mental module (as opposed to a culturally evolved by-product), small-scale hunter-gatherer societies, which best approximate ancestral conditions, lack omniscient and omnipotent supernaturals. Studies from the *Ethnographic Atlas* (1267 societies coded into 186 cultural types) show larger societies likelier to engage in external conflict and possess potent moralizing gods who punish norm violators (*17*). Modern multiculturalism and global exposure to multifarious values are increasingly challenged by fundamentalist movements to revive primary group loyalties through greater ritual commitments to ideological purity (*4*–*6*).

Religion and War

Religious beliefs involving sacred values facilitate both large-scale cooperation and enduring group conflict. For example, the mainly nonviolent Civil Rights Movement and extremely violent Al-Qaeda movement were strongly motivated by religious commitment—to "right the national sin" of slavery in one case (*18*), to "restore God's law" against indignity and injustice in the other (*19*)—however important other economic, social, and political factors. Yet, ever since the 9/11 attacks in New York, New Atheist thinkers claim that religion is chiefly responsible for war and much human misery, which its demise would greatly reduce (*15*).

In fact, explicit religious issues have motivated only a small minority of recorded wars (*20*). There is little religious cause for the internecine Russian and Chinese conflicts and world wars

Fig. 1. The backfire effect. Offering material incentives to compromise sacred values increases anger and violence toward a deal. (**A**) In a 2010 study, Iranians who regarded Iran's right to a nuclear program as a sacred value more violently opposed sacrificing Iran's nuclear program for conflict-resolution deals involving substantial economic aid, or relaxation of sanctions, than the same deals without aid or sanctions. [Photo: V. Salemi/Associated Press] (**B**) In a 2005 study in the West Bank and Gaza, Palestinian refugees who held their "right of return" to former homes in Israel as a sacred value more violently opposed abandoning this right for a Palestinian state plus substantial economic aid than the same peace deal without aid. [Photo: D. Douglas]

responsible for history's most lethal century of international conflict. Indeed, inclusive concepts of "humanity" arguably emerged with the rise of universal religions. Buddhism spread beyond India by eliminating social castes, and early Christianity became the Roman Empire's majority religion through growing social networks built on trust grounded in self-sacrificial displays (e.g., caring for non-Christians during epidemics) (*21*). Fourteenth-century historian Ibn Khaldūn found that for North African Muslim dynasties with comparable military might, long-term differences in success "have their origin in religion… group feeling [wherein] individual desires come together in agreement [and] mutual cooperation and support flourish" (*9*), the more religious societies enduring longer.

Studies with a diverse range of contemporary foragers, farmers, and herders show that professing a world religion predicts greater fairness toward ephemeral interactants (*22*). Islam's ongoing spread in sub-Saharan Africa depends on religious rituals (fasting, sobriety, charity) drawing people into tighter networks of trust and trade similar to Protestant evangelicalism's growth in the Americas and Asia (*4*, *9*).

Nonetheless, ample historical and cross-cultural evidence shows that when conflict is framed by competing religious and sacred values, intergroup violence may persist for decades, even centuries. Disputes over otherwise mundane phenomena (people, places, objects, events) then become existential struggles, as when land becomes "holy land." Secular issues become sacralized and nonnegotiable (*23*), regardless of material rewards or punishments. These values become transcendent, emotionally charged yet stable over time, and neurally processed as duties bound by rules rather than utilitarian calculations (*24*).

Sacred Values in Intractable Conflicts

Sacred values determine which social and material transactions are morally sanctioned, and can drive actions in intergroup conflict independently of material interests or consequences: as with some decisions to initiate or sustain war, when religiously motivated suicide bombers willingly die to broadcast commitment regardless of whether others are killed (*11*), or when certain Native American groups reject majority-culture offers for exploiting natural resources as violating spiritual injunctions (*25*). Studies with Palestinians, Israelis, Indonesians, Indians, Afghans, and Iranians show that offering people material incentives (large amounts of money, guarantees for a life free of political violence) to compromise sacred values can backfire, increasing violence toward compromise (*11*). Backfire effects occur both for sacred values with clear religious investment (Jerusalem, Sharia law) and those with initially none (Iran's right to nuclear capability, Palestinian refugees' right of return) (Fig. 1, A and B). While sacralization of initially secular issues blocks standard "business-like" negotiation tactics, strong symbolic gestures (sincere apologies, demonstrating respect for the other's values) generate surprising flexibility, even among militants and political leaders, and may enable subsequent material negotiations. Moreover, as with religious beliefs generally, sacred values may be reframed through novel, context-sensitive interpretation without compromising their absolute "truth" (e.g., Jerusalem is less a place than a portal to heaven, and earthly access to the portal suffices) (*10*).

We need to know more about cognitive and social mechanisms underlying sacralization of values that cement personal devotion to group norms. One hypothesis is that humans sacralize threatened group interests by incorporating them within religious ritual and rhetoric, as with Iran's nuclear program (*26*). In one experiment, Americans who were cognitively primed with reminders of religious ritual were likelier to sacralize nonreligious preferences (e.g., favor or oppose deporting illegal immigrants), measured by refusal to forsake preferences for money (*27*). In a longitudinal study, Palestinian adolescents who perceived strong threats to their people and were highly involved in religious ritual were most likely to see political issues like the right of refugees to return as absolute moral imperatives, forbidding Palestinian leaders to compromise whatever the costs (*27*).

Religion addresses problems of cooperation in large complex societies through supernatural agents that punish noncooperators, use of ritual to forge strong group bonds, and sacralization of mundane issues threatened in intergroup disputes. Future research should investigate interactions between these mechanisms. While there is no necessary link between religious belief, sacred values, and war, during intergroup conflict protagonists may transform material interests into sacred values and further consolidate them into religious beliefs. Sacred values are not exclusive to religion; mundane values may be sacralized through rituals linking them to nonreligious sacred values, like the nation. Studies could test whether religious belief reinforces sacred values by conferring divine authorship and watchfulness. The implication is that when sacred values acquire supernatural association, they become conceptually immune to reasoned challenge and subject to self-monitoring, reducing possibilities for social defection; they also become attention-arresting and memorable, rendering them culturally contagious.

Religion and sacred values inspire achievement of great virtue and vice, in spiriting folk to glory or bending will to power. They arouse strong attitudes among believers and nonbelievers, which can hinder deeper, nuanced appreciation of human conflict. In an age where religious and sacred causes are resurgent, there is urgent need for joint scientific effort to understand them. In-depth ethnography, combined with cognitive and behavioral experiments among diverse societies (including those lacking a world religion), can help identify and isolate the moral imperatives for decisions on war or peace. Neuroimaging may elucidate how religion and sacred values differ from secular beliefs and values, to better comprehend how they interact and which aspects are liable to manipulation toward violence or nonviolence. We need more developmental study on how children acquire religion and sacred values, and how people come to change or abandon them. Formal modeling of cultural evolutionary processes should join archaeological and historical efforts to mine this multidisciplinary research for insight into broad patterns of history and future prospects for human conflict and its resolution.

References and Notes

1. P. Boyer, *Nature* **455**, 1038 (2008).
2. C. Bader, *J. Sci. Study Relig.* **38**, 119 (1999).
3. A. Kay, D. Gaucher, I. McGregor, K. Nash, *Pers. Soc. Psychol. Rev.* **14**, 37 (2010).
4. P. Norris, R. Ingelhart, *Secular and Sacred* (Cambridge Univ. Press, Cambridge, 2004).
5. R. Boyd, P. Richerson, *The Origin and Evolution of Cultures* (Oxford Univ. Press, New York, 2005).
6. A. Norenzayan, A. F. Shariff, *Science* **322**, 58 (2008).
7. J. Bulbulia, R. Sosis, *Religion* **41**, 363 (2011).
8. H. Whitehouse, *Modes of Religiosity* (AltaMira, Walnut Creek, CA, 2004).
9. S. Atran, J. Henrich, *Biol. Theory* **5**, 18 (2010).
10. S. Atran, R. Axelrod, *Negotiation J.* **24**, 221 (2008).
11. J. Ginges, S. Atran, S. Sachdeva, D. Medin, *Am. Psychol.* **66**, 507 (2011).
12. D. Wilson, *Darwin's Cathedral* (Univ. of Chicago Press, Chicago, 2002).
13. J. Ginges, I. Hansen, A. Norenzayan, *Psychol. Sci.* **20**, 224 (2009).
14. P. Bloom, *Annu. Rev. Psychol.* **63**, 179 (2012).
15. R. Dawkins, D. Dennett, S. Harris, C. Hitchens, *The Four Horsemen* (video discussion, 30 September 2007); available at http://richarddawkins.net/videos/2025-the-four-horsemen-available-now-on-dvd.
16. J. Marcus, K. V. Flannery, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 18257 (2004).
17. F. Roes, M. Raymond, *Evol. Hum. Behav.* **24**, 126 (2003).
18. S. Schama, *The American Future* (HarperCollins, New York, 2009).
19. "U.S. Government Efforts to Counter Violent Extremism" (Senate Armed Services Committee, Washington, DC, 2011); available at www.jjay.cuny.edu/US_Senate_Hearing_on_Violent_Extremism.pdf.
20. C. Phillips, A. Axelrod, *Encyclopedia of Wars* (Facts on File, New York, 2007).
21. R. Stark, *The Rise of Christianity* (HarperCollins, New York, 1997).
22. J. Henrich *et al.*, *Science* **327**, 1480 (2010).
23. P. E. Tetlock, *Trends Cogn. Sci.* **7**, 320 (2003).
24. G. Berns *et al.*, *Philos Trans. R. Soc B* **367**, 754 (2012).
25. S. Atran, D. Medin, *The Native Mind and the Cultural Construction of Nature* (MIT Press, Cambridge MA, 2008).
26. M. Dehghani *et al.*, *Judgm. Decis. Mak.* **5**, 540 (2010).
27. H. Sheikh, J. Ginges, A. Coman, S. Atran, *Judgm. Decis. Mak.* **7**, 110 (2012).

Acknowledgments: We thank the U.S. Department of Defense (Air Force Office of Scientific Research, Office of Naval Research, Army Research Office) and NSF for support.

10.1126/science.1216902

HUMANCONFLICT
http://scim.ag/hum_conflict

REVIEW

ETHNICITY AND CONFLICT: THEORY AND FACTS

Joan Esteban,[1] Laura Mayoral,[1] Debraj Ray[1,2]*

Over the second half of the 20th century, conflicts within national boundaries became increasingly dominant. One-third of all countries experienced civil conflict. Many (if not most) such conflicts involved violence along ethnic lines. On the basis of recent theoretical and empirical research, we provide evidence that preexisting ethnic divisions do influence social conflict. Our analysis also points to particular channels of influence. Specifically, we show that two different measures of ethnic division—polarization and fractionalization—jointly influence conflict, the former more so when the winners enjoy a "public" prize (such as political power or religious hegemony), the latter more so when the prize is "private" (such as looted resources, government subsidies, or infrastructures). The available data appear to strongly support existing theories of intergroup conflict. Our argument also provides indirect evidence that ethnic conflicts are likely to be instrumental, rather than driven by primordial hatreds.

There are two remarkable facts about social conflict that deserve notice. First, within-country conflicts account for an enormous share of deaths and hardship in the world today. Figure 1 depicts global trends in inter- and intrastate conflict. Since the Second World War, there have been 22 interstate conflicts with more than 25 battle-related deaths per year, and 9 of them have killed at least 1000 over the entire history of conflict (*1*). The total number of attendant battle deaths in these conflicts is estimated to be around 3 to 8 million (*2*). The same period witnessed 240 civil conflicts with more than 25 battle-related deaths per year, and almost half of them killed more than 1000 (*1*). Estimates of the total number of battle deaths are in the range of 5 to 10 million (*2*). Added to the direct count of battle deaths are the 25 million noncombatant civilian (*3*) and indirect deaths due to disease and malnutrition, which have been estimated to be at least four times as high as violent deaths (*4*), as well as the forced displacements of more than 40 million individuals by 2010 (*5*). In 2010 there were 30 ongoing civil conflicts (*6*).

Second, internal conflicts often appear to be ethnic in nature. More than half of the civil conflicts recorded since the end of the Second World War have been classified as ethnic or religious (*3*, *7*). One criterion for a conflict to be classified as ethnic is that it involves a rebellion against the state on behalf of some ethnic group (*8*). Such conflicts involved 14% of the 709 ethnic groups categorized worldwide (*9*). Brubaker and. Laitin, examining the history of internal conflicts in the second half of the 20th century, are led to remark on "the eclipse of the left-right ideological axis" and the "marked ethnicization of violent challenger-incumbent contests" (*10*). Horowitz, author of a monumental treatise on the subject of ethnic conflict, observes that "[t]he Marxian concept of class as an inherited and determinative affiliation finds no support in [the] data. Marx's conception applies with far less distortion to ethnic groups.… In much of Asia and Africa, it is only modest hyperbole to assert that the Marxian prophecy has had an ethnic fulfillment" (*11*).

The frightening ubiquity of within-country conflicts, as well as their widespread ethnic nature, provokes several questions. Do "ethnic divisions" predict conflict within countries? How do we conceptualize those divisions? If it is indeed true that ethnic cleavages and conflicts are related, how do we interpret such a result? Do "primordial," ancestral ethnic hatreds trump "more rational" forms of antagonism, such as the instrumental use of ethnicity to achieve political power or economic gain? To discuss and possibly answer some of these questions is the goal of this review.

Class and Ethnicity as Drivers of Conflict

The study of human conflict is (and has been) a central topic in political science and sociology. Economics—with relatively few and largely recent exceptions—has paid little attention to the issue. [For three recent overviews, see (*12*–*14*).] Perhaps textbook economics, with its traditional respect for property rights, often presumes that the economic agents it analyzes share that respect and do not violently challenge allocations perceived to be unfair. Yet one of the notable exceptions in economics—Marx—directly or indirectly dominates the analytical landscape on conflict in the rest of the social sciences. Class struggle, or more generally, economic inequality, has been viewed as the main driver of social conflict in industrial or semi-industrial society (*15*). In Sen's words, "the relationship between inequality and rebellion is indeed a close one" (*16*).

Yet, intuitive as it might seem, this relationship doesn't receive emphatic empirical endorsement. In a detailed survey paper on the many attempts to link income inequality and social conflict empirically, Lichbach mentions 43 papers on the subject, some "best forgotten" (*17*). The evidence is thoroughly mixed, concludes Lichbach, as he cites a variety of studies to support each possible relationship between the two, and others that show no relationship at all. Midlarsky remarks on the "fairly typical finding of a weak, barely significant relationship between inequality and political violence … rarely is there a robust relationship between the two variables" (*18*).

The emphasis on economic inequality as a causal correlate of conflict seems natural, and there is little doubt that carefully implemented theory will teach us how to better read the data (see below). Yet it is worth speculating on why there is no clear-cut correlation. Certainly, economic demarcation across classes is a two-edged sword: While it breeds resentment, the very

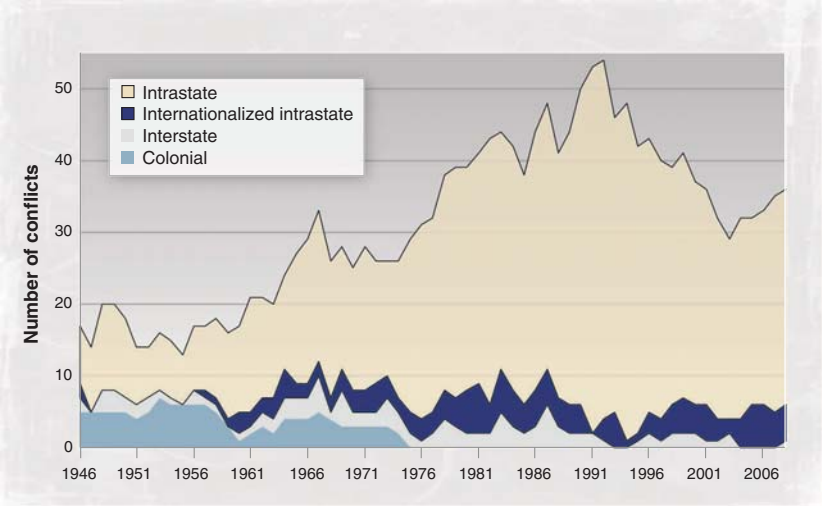


Fig. 1. Armed conflicts by type. Sources: Databased on UCDP/PRIO armed conflict database. Conflicts include cases with at least 25 battle deaths in a single year.

[1]Institut d'Anàlisi Económica, CSIC and Barcelona GSE, ES08193, Spain. [2]New York University, New York, NY 10012, USA.

*To whom correspondence should be addressed. E-mail: debraj.ray@nyu.edu

poverty of the have-nots separates them from the means for a successful insurrection. In addition, redistribution across classes is invariably an indirect and complex process.

The use of noneconomic "markers" such as ethnicity or religion addresses both these issues. Individuals on either side of the ethnic divide will be economically similar, so that the gains from such conflict are immediate: The losing group can be excluded from the sector in which it directly competes with the winners [e.g., (*11*, *19*, *20*)]. In addition, each group will have both poor and rich members, with the former supplying conflict labor and the latter supplying conflict finances (*21*). This suggests an interesting interaction between inequality and ethnicity, by which ethnic groups with a higher degree of within-group inequality will be more effective in conflict (*22*). Moreover, it has been suggested that "horizontal" inequality (i.e., inequality across ethnic groups) is an important correlate of conflict (*23*–*26*).

There are two broad views on the ethnicity-conflict nexus [e.g., (*10*, *27*)]. The "primordialist" view (*28*, *29*) takes the position that ethnic differences are ancestral, deep, and irreconcilable and therefore invariably salient. In contrast, the "instrumental" approach pioneered by (*19*) and discussed in (*10*) sees ethnicity as a strategic basis for coalitions that seek a larger share of economic or political power. Under this view, ethnicity is a device for restricting the spoils to a smaller set of individuals. Certainly, the two views interact. Exclusion is easier if ethnic groups are geographically concentrated (*30*, *31*). Strategic ethnic conflict could be exacerbated by hatreds and resentments—perhaps ancestral, perhaps owing to a recent clash of interests—that are attached to the markers themselves. Finally, under both these views, in ethnically divided societies democratic agreements are hard to reach and once reached, fragile (*32*); the government will supply fewer goods and services and redistribute less (*33*, *34*); and society will face recurrent violent conflict (*11*).

Either approach raises the fundamental question of whether there is an empirical, potentially predictive connection between ethnic divisions and conflict. To address that question, we must first define what an "ethnic division" is. Various measures of ethnic division or dominance (*35*–*37*) have been proposed. The best-known off-the-shelf measure of ethnic division is the fractionalization index, first introduced in the 1964 edition of the *Soviet Atlas Narodov Mira*, to measure ethnolinguistic fragmentation. It equals the probability that two individuals drawn at random from the society will belong to two different groups (see Box 1 for a precise definition). Ethnic fractionalization has indeed been usefully connected to per capita gross domestic product (GDP) (*38*), economic growth (*39*), or governance (*40*). But (*7*, *35*, *41*, *42*) do not succeed in finding a connection between ethnic or religious fractionalization and conflict, though it has been suggested that fractionalization appears to work better for smaller-scale conflicts, such as ethnic riots (*43*). By contrast, variables such as low GDP per capita, natural resources, environmental conditions favoring insurgency, or weak government are often statistically significant correlates of conflict (*12*, *44*). Fearon and Laitin conclude that the observed "pattern is thus inconsistent with … the common expectation that ethnic diversity is a major and direct cause of civil violence" (*7*).

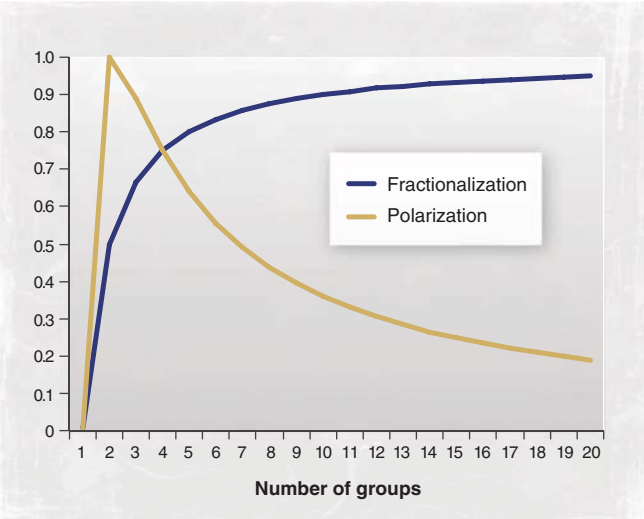


Fig. 2. Polarization, fractionalization, and the number of groups. In this illustration, all groups are of equal size, and intergroup distances are set equal to 1.

But the notion of "ethnic division" is complex and not so easily reduced to a measure of diversity. The discussion that follows will introduce a different measure—polarization—that better captures intergroup antagonism. As we shall see, polarization will be closely connected to the incidence of conflict; moreover, with a measure of polarization in place and controlled for, fractionalization, too, will matter for conflict.

Fractionalization and Polarization

As already discussed, the index of fractionalization is commonly used to describe the ethnic structure of a society (see Box 1). This index essentially reflects the degree of ethnic diversity. When groups are of equal size, the index increases with the number of groups. It reaches a maximum when everyone belongs to a different group.

When one is interested in social conflict, this measure does not seem appropriate on at least two counts. First, as social diversity increases beyond a point, intuition suggests that the likelihood of conflict would decrease rather than increase. After all, group size matters. The fact that "many are in this together" provides a sense of group identity in times of conflict. Moreover, groups need a minimum size to be credible aggressors or opponents. Second, not all groups are symmetrically positioned with respect to other groups, though the measure implicitly assumes they are. A Pushtun saying is illustrative: "Me against my brothers, me and my brothers against my cousins, me and my cousins against the world." The fractionalization measure can be interpreted as saying that every pair of groups is "equally different." Often, they are not.

Consider now the notion of polarization as introduced in (*45*–*47*). Polarization is designed to measure social "antagonism," which is assumed to be fueled by two factors: the "alienation" felt between members of different groups and the sense of "identification" with one's own group. This index is defined as the aggregation of all interpersonal antagonisms. Its key ingredients are intergroup distances (how alien groups are from each other) and group size (an indicator of the level of the group identification). Using an axiomatic approach (*45*, *48*), we obtain the specific form used in this article; see Box 1 for the precise formula.

In any society with three or more ethnic groups, the polarization measure behaves very differently from fractionalization. Unlike fractionalization, polarization declines with the continued splintering of groups and is globally maximized for a bimodal distribution of population. This is shown in Fig. 2, where groups are always of equal size and intergroup distances are equal to 1. Rather than being two different (but broadly related) ways of measuring the same thing, the two measures emphasize different aspects of a fundamentally multidimensional phenomenon. As we shall see, the differences have both conceptual and empirical bite. For instance, Montalvo and Reynal-Querol (*49*), using a simplified version of the index of polarization, show that ethnic polarization is a significant correlate of civil conflict, whereas fractionalization is not. Their contribution provides the first piece of serious econometric support for the proposition that "ethnic divisions" might affect conflict.

Despite their divergent performance in empirical work, the two measures are linked. Indeed, they are even identical if (i) group identity does not play a role and (ii) individuals feel equally alienated from members of all other groups. Which index is best to use is therefore determined by the nature of the problem at hand: on whether the sense of identity, of intergroup differentiation, or both are relevant. Group identification matters when we face problems of public import, in which the payoffs to the entire community jointly matter. Intergroup differentiation is relevant whenever the specific cultural characteristics of the other groups affect the policies that they choose, and

therefore create implications for any one group. In contrast, if social groups compete for narrow economic gains that accrue to the winners and are excludable from the losers, no opponent's victory means more or less than any other. In the theory that we outline below, these are precisely the factors that receive greatest emphasis.

Marrying Theory and Facts

A systematic econometric exploration of the links between ethnic divisions and conflict will generally take the form of a multivariate regression. The "dependent variable" we seek to explain is some measure of conflict. On the other side of the regression is our main "independent variable," which is a particular measure of "ethnic divisions," as well as a host of "control variables" that are included to capture other influences on conflict that we seek to filter out. This much is evident. The problem is (and this is true of empirical research more generally) that little discipline is often imposed on the specification of that regression. Much of that research involves the kitchen-sink approach of including all variables—usually linearly—that could possibly play a role in ethnic conflict. Such an approach is problematic on at least three counts. First, the number of plausible variables is unbounded, not just in principle but apparently also in practice: 85 different variables have been used in the literature (*50*). Trying them out in various hopeful combinations smacks uncomfortably of data-mining. Second, even if we could narrow down the set of contenders, there are many ways to specify the empirical equation that links those variables to conflict. Finally, the absence of a theory hinders the interpretation of the results.

From a statistical perspective, fractionalization and polarization are just two, seemingly equally reasonable, ways of measuring ethnic divisions. Yet they yield very different results in connecting ethnicity to conflict. Do we accept this inconsistency as yet another illustration of "measurement error"? Or is there something deeply conceptual buried here?

The results we are going to present are obtained from an explicit game-theoretic model of conflict. We then bring the predicted equilibrium of this model to data. This allows us both to test the theory and to suitably interpret the results. Perhaps the most important contribution of the theory is that it permits both polarization and fractionalization as joint drivers of conflict and explains precisely when one measure acquires more explanatory salience than the other.

We begin by presenting the recent analysis that links polarization and fractionalization to equilibrium conflict (*48*). We then describe some of the empirical findings obtained in (*51*) when confronting the predictions of the model with data.

Polarization, Fractionalization, and Conflict: Theory

A situation of open civil conflict arises when an existing social, political, or economic arrangement is challenged by an ethnic group. Whether the ethnic marker is focal for instrumental or primordial reasons is an issue that we've remarked on earlier, but at this stage it is irrelevant for our purpose. [For more on ethnic salience, see (*52*–*54*).] In such a situation, the groups involved will undertake costly actions (demonstrations, provocations, bombs, guerrilla or open warfare) to increase their probability of success. We view the aggregate of all such actions as the extent of conflict.

More precisely, suppose that there are m groups engaged in conflict. Think of two types of stakes or prizes in case of victory. One kind of prize is "public," the individual payoff from which is undiluted by one's own group size. For instance, the winning group might impose its preferred norms or culture: a religious state, the abolition of certain rights or privileges, the repression of a language, the banning of political parties, and so on. Or it might enjoy political power or the satisfaction of seeing one's own group vindicated or previous defeats avenged. Let u_{ij} be the payoff experienced by an individual member of group i in the case in which group j wins and imposes its preferred policy; we presume that $u_{ii} > u_{ij}$, which is true almost by definition. This induces a notion of "distance" across groups i and j: $d_{ij} \equiv u_{ii} - u_{ij}$, which can be interpreted as the loss to i of living under the policy implemented by j. Note that a member of group i might prefer j rather than k to be in power, and that will happen precisely when $d_{ij} < d_{ik}$.

The money-equivalent value of the public payoffs—call it π—tells us how much money individuals are ready to give up to bring the implemented policy "one unit" closer to one's own ideal policy. Its value depends in part on the extent to which the group in power can impose policies or values on the rest of society. Thus, a member of group i assigns a money value of $u_{ij}\pi$ to the ideal policy of group j.

The other type of prize is "private." Examples include the material benefits obtained from administrative or political positions, specific tax breaks, directed subsidies, bias in the allocation of public expenditure and infrastructures, access to rents from natural resources, or just plain loot. Private payoffs have two essential properties. First, group size dilutes individual benefits: The larger the group, the smaller is the return from a private prize for any one group member. Second, the identity of the winner is irrelevant to the loser since, in contrast to the "public" case, the loser is not going to extract any payoff from that fact. (If there are differential degrees of resentment over the identity of the winner, simply include this component under the public prize.) Let μ be the per capita money value of the private prize at stake.

Individuals in each group expend costly resources (time, effort, risk) to influence the probability of success. Conflict is defined to be the sum of all these resources over all individuals and all groups. The winners share the private prize and get to implement their favorite policies (the public prize). The losers have to live with the policies chosen by the winners. A conflict equilibrium describes the resulting outcome. ("Conflict equilibrium" perhaps abuses semantics to an unacceptable degree, our excuse being that we observe the game-theoretic tradition of describing the noncooperative solution to a game as a Nash "equilibrium.") It is a vector of individual actions such that each agent's behavior maximizes expected payoffs in the conflict, given the choices made by all other individuals. Note that by the word "payoff" we don't mean only some narrow monetary amount, but also noneconomic returns, such as political power or religious hegemony.

But what does the maximization of payoffs entail? Individuals are individuals, but they also have a group identity. To some extent an individual will act selfishly, and to some extent he or she will act in the interest of the ethnic group. The weight placed on the group versus the individual will depend on several factors (some idiosyncratic to the individual), but a large component will depend on the degree of group-based cohesion in the society; we return to this below. Formally, we presume that an individual places a weight of α on the total payoff of his or her group, in addition to their own payoff.

Let us measure the intensity of conflict—call it C—by the money value of the average, per capita level of resources expended in conflict. In (*48*) we argue that in equilibrium, the eventual across-group variation in the per capita resources expended has a minor effect on the aggregate level of conflict. Thus, in practice the population-normalized intensity of conflict C can be approximated well by ignoring this variation, and this simplification yields the approximate formula

$$\frac{C}{\pi + \mu} \simeq \alpha[\lambda P + (1 - \lambda)F] \qquad (1)$$

for large populations, where $\lambda \equiv \pi/(\pi + \mu)$ is the relative publicness of the prize, F is the fractionalization index, and P is a particular member of the family of polarization measures described earlier, constructed using intergroup distances d_{ij} derived from "public" payoff losses. (Box 1 describes these measures more formally and also provides a more general version of Eq. 1.) Thus, the theory tells us precisely which notions of ethnic division need to be considered. Moreover, the relationship has a particular form, which informs the empirical analysis.

This result highlights the essential role of theory for meaningful empirical work. The exogenous data of the model—individual preferences, group size, the nature and the size of the prize, and the level of group cohesion—all interact in a special way to determine equilibrium conflict intensity. The theory shows, first, that it suffices to aggregate all the information on preferences and

group sizes into just two indices—*F* and *P*—capturing different aspects of the ethnic composition of a country. Second, the weights on the two distributional measures depend on the composition of the prize and on the level of group commitment. In particular, the publicness of the prize (reflected in a high value of λ) reinforces the effect of polarization, whereas high privateness of the prize (low λ) reinforces the effect of fractionalization. Not surprisingly, high group cohesion α enhances the effect of both measures on conflict.

The publicness of the prize is naturally connected to both identification and alienation—and therefore to polarization. With public payoffs, group size counts twice: once, because the payoffs accrue to a larger number, and again, because a larger number of individuals internalize that accrual and therefore contribute more to the conflict. Intergroup distances matter, too: The precise policies interpreted by the eventual winner continue to be a cause of concern for the loser. Both these features—the "double emphasis" on group size and the use of distances—are captured by the polarization measure *P*; see Box 1 for more details. By contrast, when groups fight for a private payoff—say money—one winner is as bad as another as long as my group doesn't win, and measures based on differences in intergroup alienation become useless. Moreover, with private payoffs, group identification counts for less than it does with public payoffs, as group size erodes the per capita gain from the prize. The resulting index that is connected to this scenario is one of fractionalization (see Box 1).

In short, the theory tells us to obtain data on *P* and *F* and combine them in a particular way. It tells us that when available, we should attempt to obtain society-level data for group cohesion α and relative publicness λ and enter them in the way prescribed by Eq. 1. With this in mind, we now bring the theory to the data.

Taking the Theory to Data

We study 138 countries over 1960 to 2008, with the time period divided into 5-year intervals. That yields a total of 1125 observations (in most cases). Some of the variables in the theory are not directly observable, and so we will use proxies. For a complete set of results, see (*51*) and the accompanying Web Appendix.

We measure conflict intensity in two ways. The first is the death toll. Using data from the jointly maintained database under the Uppsala Conflict Data Program and the Peace Research Institute of Oslo (UCDP/PRIO) (*1*), we construct a discrete measure of conflict—PRIO-C—for every 5-year period and every country as follows: PRIO-C is equal to 0 if the country is at peace in those 5 years; to 1 if it has experienced low-intensity conflict (more than 25 battle-related deaths but less than 1000) in any of these years; or to 2 if the country has been in high-level conflict (more than 1000 casualties) in any of the 5 years. Despite the overall popularity of UCDP/PRIO, this is an admittedly coarse measure of deaths, based on only three categories (peace, low conflict, and high conflict) defined according to ad hoc thresholds, and it reports conflicts only when one of the involved parties is the state. To overcome these two problems, we use a second measure of intensity: the *Index of Social Conflict* (ISC) computed by the Cross-National Time-Series Data Archive (*55*). It provides a continuous measure of several manifestations of social unrest, with no threshold dividing "peace" from "war." The index ISC is formed by taking a weighted average over eight different manifestations of internal conflict, such as politically motivated assassinations, riots, guerrilla warfare, etc.

Our core independent variables are the indices *F* and *P*. To compute these indices, we need the population size of different ethnic groups for every country and a proxy for intergroup distances. For demographic information on groups, we use the data set provided by (*9*), which identifies over 800 "ethnic and ethno-religious" groups in 160 countries. For intergroup distances, we follow (*9*, *56*, *57*) and use the linguistic distance between two groups as a proxy for group "cultural" distances in the space of public policy.

Linguistic distance is defined on a universal language tree that captures the genealogy of all languages (*58*). All Indo-European languages, for instance, will belong to a common subtree. Subsequent splits create further "sub-subtrees," down to the current language map. For instance, Spanish and Basque diverge at the first branch, since they come from structurally unrelated language families. By contrast, the Spanish and Catalan branches share their first seven nodes: Indo-European, Italic, Romance, Italo-Western, Western, Gallo-Iberian, and Ibero-Romance languages. We measure the distance between two languages as a function of the number of steps we must retrace to find a common node. The results are robust to alternative ways of mapping linguistic differences into distances.

Linguistic divisions arise because of population splits. Languages with very different origins reveal a history of separation of populations going back several thousand years. For instance, the separation between Indo-European languages and all others occurred around 9000 years ago (*59*). In contrast, finer divisions, such as those between Spanish and Catalan, tend to be the result of more recent splits, implying a longer history of common evolution. Consistent with this view, there is evidence showing a link between the major language families and the main human genetic clusters (*60*, *61*).

The implicit theory behind our formulation is that linguistic distance is associated with cultural distance, stemming from the chronological relation of language trees to group splittings and, therefore, to independent cultural (and even genetic) evolution. That argument, while obviously not self-evident, reflects a common trade-off. The disadvantage is obvious: Linguistic distances are at best an imperfect proxy for the unobserved "true distances." But something closer to the unobserved truth—say, answers to survey questions about the degree of intergroup antagonism, or perhaps a history of conflict—have the profound drawback of being themselves affected by the very outcomes they seek to explain, or being commonly driven (along with the outcome of interest) by some other omitted variable. That is, such variables are endogenous to the problem at hand. The great advantage of linguistic distances is that a similar charge cannot be easily leveled against them. Whether the trade-off is made well here is something that a mixture of good intuition and final results must judge.

In our specifications, we also control for other variables that have been shown to be relevant in explaining civil conflict (*12*): population size (POP), because conflict is population-normalized in the theory; gross domestic product per capita (GDPPC), which raises the opportunity cost of supplying conflict resources; natural resources (NR), measured by the presence of oil or diamonds, which affects the total prize; the percentage of mountainous terrain (MOUNT), which facilitates guerrilla warfare; noncontiguity (NCONT), referring to countries with territory separated from the land area containing the capital city either by another territory or by 100 km of water; measures of the extent of democracy (DEMOC); the degree of power (PUB) afforded to those who run the country, which is a proxy for the size of the public prize (more on this below); time dummies to capture possible global trends; and regional dummies to capture patterns affecting entire world regions. Finally, because current conflict is deeply affected by past conflict, we use lagged conflict as an additional control in all our specifications.

Our exercise implements Eq. 1 in three ways. First, we run a cross-sectional regression of conflict on the two measures of ethnic division. Second, we independently compute a degree of relative publicness of payoffs for each country and include this in the regression. Third, we add separate proxies of group cohesion for all the countries. Each of these steps takes us progressively closer to the full power of Eq. 1, but with the potential drawback that we need proxies for an increasing number of variables.

To form a relative publicness index by country, we proxy π and μ for every country. Begin with a proxy for the private payoff μ. It seems natural to associate μ with rents that are easily appropriable. Because appropriability is closely connected to the presence of resources, we approximate the degree of "privateness" in the prize by asking if the country is rich in natural resources. Typically, oil and diamonds are the two commodities most frequently associated with the

"resource curse" (*62*, *63*). Data on the quantity of diamonds produced is available (*64*), but information on quality (and associated price) is scarce, making it very difficult to estimate the monetary value of diamond production. Diamond prices per carat can vary by a factor of 8 or more, from industrial diamonds ($25 a carat in 2001) to high-quality gemstones ($215 per carat in 2001) (*63*). Hence, we focus exclusively on oil in this exercise. We use the value of oil reserves per capita, OILRSVPC, as a proxy for μ.

Next, we create an index of "publicness," PUB, by measuring the degree of power afforded to those who run the country, "more democratic" being regarded as correlated with "less power" and consequently a lower valuation of the public payoff to conflict. We use four different proxies to construct the index: (i) the lack of executive constraints, (ii) the level of autocracy, (iii) the degree to which political rights are flouted, and (iv) the extent of suppression of civil liberties. We use time-invariant dummies of these variables based on averages over the sample, because short-run changes are likely to be correlated with the incidence of conflict.

Our proxy for the relative publicness of the prize is given by

$$\Lambda \equiv (\gamma \mathrm{PUB} \times \mathrm{GDPPC}) / (\gamma \mathrm{PUB} \times \mathrm{GDPPC} + \mathrm{OILRSVPC}) \quad (2)$$

where we multiply the PUB indicator by per capita GDP to convert the "poor governance" variables into monetary equivalents. The "conversion factor" γ makes the privateness and publicness variables comparable and allows us to combine them to arrive at the ratio Λ. In the empirical exercise we present here, we set γ equal to 1. But the results are robust to the precise choice of this parameter; see the Web Appendix to (*51*).

Finally, we proxy the level of group cohesion α by exploiting the answers to a set of questions in the 2005 wave of the World Values Survey (*65*). We use the latest wave available because it covers the largest number of countries. One could argue that the answers might be conditioned by the existence of previous or contemporary conflict. Hence, the questions we have selected do not ask about commitment to specific groups but address issues like adherence to social norms, identification with the local community, the importance of helping others, and so on. We compute the country average of individual scores on this set of questions and denote this by A; see (*51*) for a list of the questions.

What the Data Say

As already mentioned, we proceed in three steps. First, we examine the strength of the cross-country relationship between conflict intensity and the two indices of ethnic division, with all controls in place, including time and regional dummies. The estimated coefficients will address the importance of the two independent variables as determinants of conflict intensity. In the second stage, we step closer to the full model and interact the distributional indices with country-specific measures of the relative publicness λ of payoffs, just as in Eq. 1. Finally, we test the full model by adding to the previous specification the extent of group cohesion α independently computed for each country. In both the second and third stages, we also retain the two distributional indices without interaction to verify whether the significance comes purely from the ethnic structure of the different countries or because this structure interacts with λ and α in the way predicted by the theory.

In stage 1, then, we regress conflict linearly on the two distributional indices and all other controls. Columns 1 and 2 in Table 1 record the results for each specification of the conflict intensity variable—PRIO-C and ISC. Ethnicity turns out to be a significant correlate of conflict, in sharp contrast to the findings of the previous studies mentioned above. Throughout, P is highly significant and positively related to conflict. F also has a positive and significant coefficient.

Apart from statistical significance, the effect of these variables is quantitatively important. Taking column 1 as reference, if we move from the median polarized country (Germany) to the country in the 90th percentile of polarization (Niger), while changing no other institutional or economic variable in the process and evaluating

Box 1. A model of conflict and distribution.

The two measures of ethnic divisions discussed in this article are both based on the same underlying parameters: the number of groups m and total population N, the population N_i of each group, and the intergroup distances d_{ij}. Polarization and fractionalization are given by

$$P = \sum_{i=1}^{m} \sum_{j=1}^{m} n_i^2 n_j d_{ij} \text{ and } F = \sum_{i=1}^{m} \sum_{j \neq 1}^{m} n_i n_j$$

where $n_i = N_i/N$ is the population share of group i. The distinction between P and F is superficial at first sight but is of great conceptual importance. The squaring of population shares in P means that group size matters over and above the mere counting of individual heads implicit in F. In addition, fractionalization F discards intergroup distances and replaces them with 0 or 1 variable.

The theory developed in (*48*) and summarized below links these measures to conflict incidence. There are m groups engaged in conflict. The winner enjoys two sorts of prizes: One is "private" and the other is "public." Let μ be the per capita value of the private prize at stake. Let u_{ij} be the utility to an individual member of group i from the policy implemented by group j. For any i the utility from the ideal policy is strictly higher than any other policy; that is, $u_{ii} > u_{ij}$. Then, the "distance" between i and j is $d_{ij} \equiv u_{ii} - u_{ij}$, so that the loss to i from j's ideal policy is d_{ij}. Let π be the amount of money an individual is willing to give up in order to bring the implemented policy one unit toward her ideal policy. Then, we can say that the monetary value to a member of group i of policy j is πu_{ij} and the loss relative to the ideal policy is πd_{ij}. Individuals in each group expend resources r to influence the probability of success of their own group. Write the income equivalent cost to such expenditure as $c(r)$ and assume that c is increasing, smooth, and strictly convex, with $c'(0) = 0$. Add individual contributions in group i to obtain group contribution R_i. Assume that the probability of success for group i is given by $p_i = R_i/R_N$, where $R_N \equiv \sum_i R_i$. Measure conflict intensity in population-normalized form by $\rho = R_N/N$.

The direct payoff to a person in group i who expends resources r is given by $\pi\mu_{ii} + p_i\mu\,/n_i - \sum_{j=1}^{m} p_j \pi d_{ij} - c(r)$. Individuals also care about the payoff to the other group members. When deciding on how much r to contribute, individuals seek to maximize the sum of their direct payoff and the total of the other group members, weighted by a group commitment factor α. Note that the optimal contribution r_i by a member of group i depends on the contributions made by all other individuals. We focus on the Nash equilibrium of this strategic game: the vector of actions with the property that all are the best response to each other. We prove that such an equilibrium always exists and that it is unique.

Note now that $c'(r)$ is the implicit "price" in sacrificed income that an individual is willing to pay for an extra unit of effort contributed to conflict. We then define the per capita normalized intensity of conflict C as the value of the resources expended, $C = c'(\rho)\rho$. Hence, $[C/(\pi + \mu)]$ is the ratio of the resources wasted in conflict relative to the stakes, all expressed in monetary terms. Proposition 2 in (*48*) shows that the equilibrium intensity of conflict C is approximately determined as follows:

$$\frac{C}{\pi+\mu} \simeq \alpha[\lambda P + (1-\lambda)F] + \lambda(1-\alpha)\frac{G}{N} + \frac{(1-\lambda)(1-\alpha)(m-1)}{N}$$

where $\lambda \equiv \pi/(\pi + \mu)$ is the relative publicness of the prize, and where G is a third measure of ethnic distribution, the Greenberg-Gini index: $G = \sum_{i=1}^{m}\sum_{j=1}^{m} n_i n_j d_{ij}$. Its influence wanes with population size, and we've ignored it in this essay, though (*48*, *51*) contain a detailed discussion of all three measures.

For large populations, the expression above reduces to the one in the main text.

those variables at their means, the predicted probability of experiencing conflict (i.e, the probability of observing strictly positive values of PRIO-C) rises from ~16 to 27%, which implies an increase of 69%. Performing the same exercise for F (countries at the median and at the 90th percentile of F are Morocco and Cameroon, respectively) takes us from 0.19 to 0.25% (an increase of 32%). These are remarkably strong effects, not least because in the thought experiment we change only the level of polarization or fractionalization, keeping all other variables the same.

Figure 3 depicts two world maps. The dots in each map show the maximum yearly conflict intensity experienced by each country; smaller dots meet the 25-death PRIO criterion, whereas larger dots satisfy the 1000-death criterion. Although these maps cannot replicate the deeper findings of the statistical analysis, they clearly show the positive relationship between conflict and ethnic divisions.

In stage 2, we consider the cross-country variation in relative publicness; recall our proxy index Λ from (*2*). In columns 3 and 4 in Table 1, the main independent variables are $P^*\Lambda$ and $F^*(1-\Lambda)$, just as specified by the theory; see Eq. 1. This allows us to test whether the interacted indices of ethnic fractionalization and polarization are significant. We also include the noninteracted indices to examine whether their significance truly comes from the interaction term. Indeed, polarization interacted with Λ is positive and highly significant, and the same is true of fractionalization interacted with $1-\Lambda$. These results confirm the relevance of both polarization and fractionalization in predicting conflict once the variables are interacted with relative publicness in the way suggested by the theory.

Table 1. Ethnicity and conflict. All specifications use region and time dummies, not shown explicitly. *P* values are reported in parentheses. Robust standard errors adjusted for clustering have been used to compute *z* statistics. Columns 1, 3, and 5 are estimated by maximum likelihood in an ordered logit specification, and columns 2, 4, and 6 by OLS. GDPPC: log of gross domestic product per capita; POP: log of population; NR: a dummy for oil and/or diamonds in columns 1 and 2 and oil reserves per capita (OILRSVPC) in columns 3 to 6; MOUNT: percentage of mountainous territory; NCONT: noncontiguous territory (see text); POLITICS is DEMOC in columns 1 and 2, and the index PUB times GDPPC (the numerator of λ) for the remaining columns; LAG, lagged conflict in previous 5-year interval; CONST, constant term.

Variable	1 PRIO-C	2 ISC	3 PRIO-C	4 ISC	5 PRIO-C	6 ISC
P	***5.16	***19.50	−1.48	−16.33	−1.47	−23.80
	(0.001)	(0.002)	(0.606)	(0.227)	(0.701)	(0.212)
F	*0.93	*3.56	0.76	0.31	0.87	−0.16
	(0.070)	(0.061)	(0.196)	(0.878)	(0.403)	(0.710)
$P\Lambda$			***11.174	***61.89		
			(0.003)	(0.001)		
$F(1-\Lambda)$			*1.19	***10.40		
			(0.097)	(0.000)		
$P\Lambda$ A					*12.65	***90.32
					(0.087)	(0.010)
$F(1-\Lambda)A$					2.54	**13.15
					(0.164)	(0.018)
GDPPC	**−0.34	***−2.26	*−0.36	***− 3.02	−0.25	***−3.68
	(0.047)	(0.004)	(0.080)	(0.001)	(0.375)	(0.007)
POP	***0.24	***1.14	***0.21	***1.30	*0.09	**1.29
	(0.000)	(0.000)	(0.001)	(0.000)	(0.166)	(0.013)
NR	−0.27	−0.53	−0.00	0.00	**0.00	*0.00
	(0.178)	(0.497)	(0.570)	(0.432)	(0.011)	(0.090)
MOUNT	0.00	0.02	0.00	*0.03	*0.01	**0.05
	(0.537)	(0.186)	(0.362)	(0.061)	(0.060)	(0.020)
NCONT	***1.06	***4.55	**0.77	***4.28	***1.37	***5.89
	(0.001)	(0.001)	(0.026)	(0.001)	(0.004)	(0.000)
POLITICS	0.18	0.29	−0.00	**−0.00	0.00	−0.00
	(0.498)	(0.789)	(0.328)	(0.026)	(0.886)	(0.374)
LAG	***1.99	***0.46	***1.94	***0.44	***1.84	***0.40
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
CONST	–	0.90	–	9.19	–	15.40
		(0.915)		(0.398)		(0.328)
(Pseudo)-R^2	0.35	0.43	0.36	0.44	0.40	0.43
Observations	1125	1111	1104	1090	447	443
Countries	138	138	138	138	53	53

It is of interest that the level terms P and F are now no longer significant. Indeed, assuming that our proxy for relative publicness accurately captures all these issues at stake, this is precisely what the model would predict. For instance, polarization should have no further effect over and beyond the "λ-channel": Its influence should dip to zero when there are no public goods at stake. That our estimate Λ happens to generate exactly this outcome is of interest. But the public component of that estimate is built solely on the basis of governance variables. If this eliminates all extraneous effects of polarization (as it indeed appears to do), it could suggest that primordial factors such as pure ethnic differences per se have little to do with ethnic conflict.

Finally, in our third stage, we allow group cohesion to vary across countries. Unfortunately, we are able to proxy A for just 53 countries, and this restricts the number of our observations to 447. Columns 5 and 6 of Table 1 examine this variant. In this specification, the independent variables are exactly in line with those described by the model, though we've had to sacrifice data. We use precisely the combinations asked for by the theory: polarization is weighted both by Λ and by A, and fractionalization by $(1-\Lambda)$ and by A again. We continue to use the direct terms P and F, as well as the controls. The results continue to be striking. The composite terms for polarization are significant, whereas the levels are not. The composite term for fractionalization is highly significant when we focus on smaller-scale social unrest, as measured by ISC, but it is marginally nonsignificant in column 5. The level terms of F continue to be insignificant. This behavior of fractionalization mirrors previous results that showed the nonrobust association of F and different manifestations of conflict (*7*, *35*).

What Have We Learned?

Existing ethnographic literature makes it clear that most within-country social conflicts have a strong ethnic or religious component. But the ubiquity of ethnic conflict is a different proposition from the assertion of an empirical link between existing ethnic divisions and conflict intensity. We've argued in this article that such a link can indeed be unearthed, provided that we're willing to write down a theory that tells us what the appropriate notion of an "ethnic division" is. The theory we discuss points to one particular measure—polarization—when the conflict is over public payoffs such as political power. It also points to a different measure—fractionalization—when the conflict is over private payoffs such as access to resource rents. Indeed, the theory also tells us how to combine the measures when there are elements of both publicness and privateness in the prize. With these considerations in mind, the empirical links between ethnicity and conflict are significant and strong.

The theory and empirical strategy together allow us to draw additional interesting inferences.

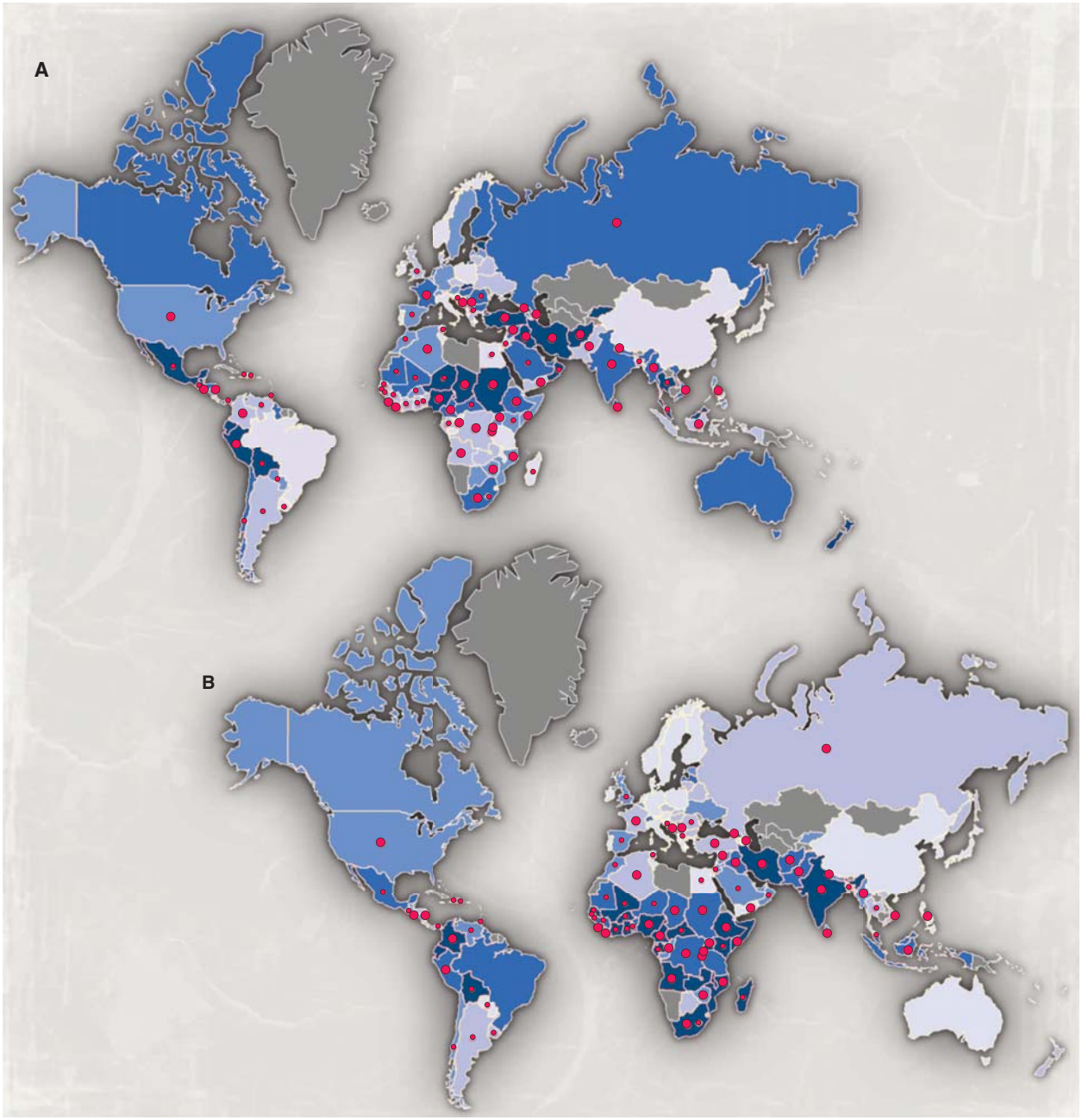


Fig. 3. Ethnicity and conflict. Dots represent the maximum yearly conflict intensity that each country has experienced over the period; smaller dots meet the 25-death PRIO criterion, whereas larger dots satisfy the 1000-death criterion. Darker colors signify higher degrees of polarization (**A**) or fractionalization (**B**). Countries for which no data are available are depicted in gray.

First, we find conclusive evidence that civil conflict is associated with (and possibly driven by) public payoffs, such as political power, and not just by the quest for private payoffs or monetary gain. Otherwise only fractionalization would matter, and not polarization. Second, the disappearance of the level effects of P and F once interactions with relative publicness are introduced (as specified by the theory) strongly suggests that ethnicity matters, not intrinsically as the primordialists would claim, but rather instrumentally, when ethnic markers are used as a means of restricting political power or economic benefits to a subset of the population.

One might object that the results are driven by the peculiarities of some regions that exhibit both highly polarized ethnicities and frequent and intense conflicts. Africa is a natural candidate that comes to mind. However, if we use regional controls or repeat the exercise by removing one continent at a time from the data set, we obtain exactly the same results (*51*).

It is too much to assert that every conflict in our data set is ethnic in nature and that our ethnic variables describe them fully. Consider, for instance, China or Haiti or undivided Korea, which have experienced conflict and yet have low polarization and fractionalization. All conflict is surely not ethnic, but what is remarkable is that so many of them are, and that the ethnic characteristics of countries are so strongly connected with the likelihood of conflict. Yet we must end by calling for a deeper exploration of the links between economics, ethnicity, and conflict.

This paper takes a step toward the establishment of a strong empirical relationship between

conflict and certain indicators of ethnic group distribution, one that is firmly grounded in theory. In no case did we use income-based groups or income-based measures, and in this sense our study is perfectly orthogonal to those that attempt to find a relationship between economic inequality and conflict, such as those surveyed in (*17*). Might that elusive empirical project benefit from theoretical discipline as well, just as the ethnicity exercise here appears to? It well might, and such an endeavor should be part of the research agenda. But with ethnicity and economics jointly in the picture, it is no longer a question of one or the other as far as empirical analysis is concerned. The interaction between these two themes now takes center stage. As we have already argued, there is a real possibility that the economics of conflict finds expression across groups that are demarcated on other grounds: religion, caste, geography, or language. Such markers can profitably be exploited for economic and political ends, even when the markers themselves have nothing to do with economics. A study of this requires an extension of the theory to include the economic characteristics of ethnic groups and how such characteristics influence the supply of resources to conflict. It also requires the gathering of group data at a finer level that we do not currently possess. In short, a more nuanced study of the relative importance of economic versus primordial antagonisms must await future research.

References and Notes

1. N. P. Gleditsch, P. Wallensteen, M. Eriksson, M. Sollenberg, H. Strand, *J. Peace Res.* **39**, 615 (2002).
2. B. Lacina, N. P. Gleditsch, *Eur. J. Popul.* **21**, 145 (2005).
3. *Political Instability Task Force*, http://globalpolicy.gmu.edu/pitf/pitfcode.htm; data accessed March 2012.
4. Global Burden of Armed Violence, *Geneva Declaration* (2008), http://www.genevadeclaration.org/; data accessed March 2012.
5. UN Web site, www.unocha.org; data accessed March 2012.
6. *Upsala Conflict Data Program*, www.pcr.uu.se/research/UCDP/; data accessed March 2012.
7. J. Fearon, D. Laitin, *Am. Polit. Sci. Rev.* **97**, 75 (2003).
8. J. D. Fearon, in *The Oxford Handbook of Political Economy*, B. R. Weingast, D. A. Wittman, Eds. (Oxford Univ. Press, Oxford, 2006), p. 852.
9. J. D. Fearon, *J. Econ. Growth* **8**, 195 (2003).
10. R. Brubaker, D. D. Laitin, *Annu. Rev. Sociol.* **24**, 423 (1998).
11. D. Horowitz, *Ethnic Groups in Conflict* (Univ. of California Press, Berkeley, CA, 1985).
12. C. Blattman, E. Miguel, *J. Econ. Lit.* **48**, 3 (2010).
13. M. Garfinkel, S. Skaperdas, Eds., *Oxford Handbook of the Economics of Peace and Conflict* (Oxford Univ. Press, Oxford, 2012).
14. M. Garfinkel, S. Skaperdas, in *Handbook of Defense Economics*, T. Sandler, K. Hartley, Eds. (North Holland, Amsterdam, 2007), vol. 2.
15. R. Dahrendorf, *Class and Class Conflict in Industrial Society* (Stanford Univ. Press, Stanford, CA, 1959).
16. A. Sen, *On Economic Inequality* (Clarendon Press, Oxford, 1973).
17. M. I. Lichbach, *World Polit.* **41**, 431 (1989).
18. M. I. Midlarsky, *Am. Polit. Sci. Rev.* **82**, 491 (1988).
19. R. H. Bates, in *State Versus Ethnic Claims: African Policy Dilemmas*, D. Rothchild, V. A. Olorunsola, Eds. (Westview, Boulder, CO, 1983).
20. A. Varshney, *Ethnic Conflict and Civic Life: Hindus and Muslims in India* (Yale Univ. Press, New Haven, CT, 2002).
21. J. Esteban, D. Ray, *Am. Econ. Rev.* **98**, 2185 (2008).
22. J. Esteban, D. Ray, *J. Eur. Econ. Assoc.* **9**, 496 (2011).
23. N. Sambanis, "Theory of Civil War," in *Understanding Civil War: Evidence and Analysis*, vol. 1, *Africa*, P. Collier, N. Sambanis, Eds. (World Bank, Washington, DC, 2005).
24. F. Stewart, "Horizontal inequalities: A neglected dimension of development," Helsinki, UNU-WIDER Working Paper, www.wider.unu.edu/publications/annual-lectures/en_GB/AL5/ (2002).
25. L. E. Cederman, N. B. Weidmann, K. S. Gleditsch, *Am. Polit. Sci. Rev.* **105**, 478 (2011).
26. G. Østby, *J. Peace Res.* **45**, 143 (2008).
27. J. Fearon, in *Oxford Handbook of Political Economy*, B. R. Weingast, D. Wittman, Eds. (Oxford Univ. Press, Oxford, 2006).
28. S. Huntington, *The Clash of Civilizations and the Remaking of World Order* (Simon & Schuster, New York, 1996).
29. M. Ignatieff, *Blood and Belonging* (Noonday, London, 1993).
30. J. Matuszeski, F. Schneider, "Patterns of ethnic group segregation and civil conflict"; www.cgdev.org/doc/events/02.09.07/Matuszeski-JMP.pdf (2006).
31. N. B. Weidmann, J. K. Rød, L.-E. Cederman, *J. Peace Res.* **47**, 491 (2010).
32. A. Rabushka, K. A. Shepsle, *Politics in Plural Societies: A Theory of Democratic Instability* (Merrill, Columbus, OH, 1972).
33. A. Alesina, A. Devleeschauwer, W. Easterly, S. Kurlat, R. Wacziarg, *J. Econ. Growth* **8**, 155 (2003).
34. A. Alesina, E. La Ferrara, *J. Econ. Lit.* **43**, 762 (2005).
35. P. Collier, A. Hoeffler, *Oxf. Econ. Pap.* **56**, 563 (2004).
36. K. Desmet, I. Ortuño-Ortín, S. Weber, *J. Eur. Econ. Assoc.* **7**, 1291 (2009).
37. L.-E. Cederman, L. Girardin, *Am. Polit. Sci. Rev.* **101**, 173 (2007).
38. A. Alesina, R. Baqir, W. Easterly, *Q. J. Econ.* **114**, 1243 (1999).
39. W. Easterly, R. Levine, *Q. J. Econ.* **112**, 1203 (1997).
40. P. Mauro, *Q. J. Econ.* **110**, 681 (1995).
41. P. Collier, A. Hoeffler, *Oxf. Econ. Pap.* **50**, 563 (1998).
42. N. Sambanis, *J. Conflict Resolut.* **48**, 814 (2004).
43. C. Tilly, *Politics of Collective Violence* (Cambridge Univ. Press, Cambridge, 2003).
44. E. Miguel, S. Satyanath, E. Sergenti, *J. Polit. Econ.* **112**, 725 (2004).
45. J. Esteban, D. Ray, *Econometrica* **62**, 819 (1994).
46. J.-Y. Duclos, J. Esteban, D. Ray, *Econometrica* **72**, 1737 (2004).
47. M. C. Wolfson, *Am. Econ. Rev.* **84**, 353 (1994).
48. J. Esteban, D. Ray, *Am. Econ. Rev.* **101**, 1345 (2011).
49. J. Montalvo, M. Reynal-Querol, *Am. Econ. Rev.* **95**, 796 (2005).
50. H. Hegre, N. Sambanis, *J. Conflict Resolut.* **50**, 508 (2006).
51. J. Esteban, L. Mayoral, D. Ray, *Am. Econ. Rev.* **102**, in press, with a Web Appendix located at www.econ.nyu.edu/user/debraj/Papers/EstebanMayoralRayAERApp.pdf (2012).
52. F. Caselli, W. J. Coleman, "On the theory of ethnic conflict"; http://personal.lse.ac.uk/casellif.
53. J. Robinson, *Econ. of Gov.* **2**, 85 (2001).
54. J. Esteban, D. Ray, *J. Econ. Theor.* **87**, 379 (1999).
55. A. Banks, Databanks International, Jerusalem (2008); http://personal.lse.ac.uk/casellif/papers/ethnic.pdf.
56. D. D. Laitin, *Am. J. Pol. Sci.* **44**, 142 (2000).
57. K. Desmet, I. Ortuño-Ortín, R. Wacziarg, *J. Dev. Econ.* **97**, 322 (2012).
58. Ethnologue, www.ethnologue.com.
59. R. D. Gray, Q. D. Atkinson, *Nature* **426**, 435 (2003).
60. L. L. Cavalli-Sforza, A. Piazza, P. Menozzi, J. Mountain, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6002 (1988).
61. E. M. S. Belle, G. Barbujani, *Am. J. Phys. Anthropol.* **133**, 1137 (2007).
62. M. Ross, *J. Peace Res.* **41**, 337 (2004).
63. M. Ross, *Annu. Rev. Polit. Sci.* **9**, 265 (2006).
64. M. Humphreys, *J. Conflict Resolut.* **49**, 508 (2005).
65. www.worldvaluessurvey.org.

Acknowledgments: We gratefully acknowledge financial support from Ministerio de Economía y Competitividad project ECO2011-25293 and from Recercaixa. J.E. and L.M. acknowledge financial support from the AXA Research Fund. The research of D.R. was funded by NSF grant SES-0962124. We thank two referees for valuable comments and are grateful to R. Jayaraman for suggestions that improved the manuscript.

10.1126/science.1222240

REVIEW

MODELING ARMED CONFLICTS

Moshe Kress

Armed conflicts have been prevalent throughout history, in some cases having very great consequences. To win, one needs to understand the characteristics of an armed conflict and be prepared with resources and capabilities for responding to its specific challenges. An important tool for understanding these characteristics and challenges is a model—an abstraction of the field of conflict. Models have evolved through the years, addressing different conflict scenarios with varying techniques.

Armed conflicts start because people disagree. They disagree on controlling territory, economic interests (such as natural resources), religion, culture, and ideology. Disagreements may lead to tense disputes, which can result in armed conflicts of various scales. History is cluttered with armed conflicts involving tribes, states, insurgencies, guerrilla groups, and terrorist organizations. Attaining victory is associated with questions regarding the use of resources, strategy, and tactics. Answering these questions is not easy; it requires a clear definition of objectives, a detailed review of capabilities and constraints, and a careful analysis of possible scenarios and courses of action. For this, military and defense analysts use models, which are abstractions of armed conflicts, their environments, and their possible realizations. These models, henceforth called armed conflict (AC) models, are used for analyzing threat situations, military operations, and force structures.

Operations Research, Naval Postgraduate School, Monterey, CA 93943, USA. E-mail: mkress@nps.edu

Armed conflicts encompass hard factors such as weapons, personnel, logistics, communication systems, and sensors, as well as soft factors such as training, tactics, leadership, situational awareness, and coordination. The "art" of AC modeling is to choose those few dominant factors that make a model manageable, meaningful, and useful.

AC models have been useful for studying cause-and-effect relations on the battlefield and to help choose among alternate systems and courses of action. AC models provide insights regarding defense issues such as the relative importance of weapons and sensors, the effect of certain tactics (such as the concentration of forces), the impact of a new weapon system, etc. For this purpose, there exist reliable and complete data sets such as estimates of hit probabilities and ranges of weapons, detection rates of sensors, capabilities of command–and-control systems, and other physical characteristics of weapons and equipment. These data are collected from tests, field experiments, simulations, and, on some rare occasions, combat events. However, although some small-scale and physical (such as firing or surveillance) models can provide good predictions of specific battlefield outcomes based on the aforementioned data, in general AC models are affected by limited and unreliable operations data, rapid changes in the environment and structure of military campaigns, and the enhanced uncertainty introduced by human behavior in the stressful combat environment. Although limited in projection power, campaign models, which encompass the combined effects of large-scale joint operations, are important for comparing alternatives and improving defense planning processes by highlighting crucial aspects. In addition to supporting decision-making, AC models are also used for educating and training military officers and defense executives. In particular, recent advances in computer technologies facilitate generating virtual environments for training combatants at the tactical level (*1*).

Fig. 1. (**A**) A U.S. Marine Corps physical simulation in a sand table in the desert in Yuma, Arizona. (**B**) Monitoring of airspace during a joint and binational exercise executed through the Air Force's Northern Distributed Mission Operations (DMO) system. DMO is a system of interconnected simulators that allows warfighters from across the United States and Canada to participate simultaneously in the same live scenario. [U.S. Air Force Photo by Angela Pope]

Historical and Classical Models

Probably the earliest AC models in the modern era were Kriegsspiels—war games developed in the early 19th century for training, planning, and testing military operations in the Prussian Army. A game was played on a table with modular combinations of precast terrain formation, gaming pieces, and dice, and it progressed following rules based on actual military maneuvers and battles. These models played an important role in the military operations that led to German unification in 1871 (*2*). Up until World War II (WWII), war games were the only AC models, used primarily for relatively small-scale (for example, battalion-sized) battles. The big leap in AC modeling occurred during WWII, when more scientific approaches to military modeling set the foundation for a newly emerging scientific field—operations research (*3*).

During WWII, a simple AC model changed the way supply convoys were dispatched across the Atlantic to provide the lifeblood for the Allied forces (*4*), and a back-of-the envelope calculation improved the tactics for hunting German submarines (*3*). More recent examples are mentioned below. There are many other examples of decisions supported by AC models that remain classified. With the advent of computing technologies, combat models have become more complex and detailed, requiring elaborate preparations that include parameter inputs and extensive postprocessing work and analysis. Many of these models are high-resolution large-scale simulations or agent-based models in which the conflict is modeled by a group of individual entities (*5*). Such models are used by the defense establishment to help choose weapon systems, such as air defense (*6*); select multibillion-dollar defense projects; determine operational plans (*7*); and set long-term goals, as in the Quadrennial Defense Review.

The validity of AC models is typically assessed by the defense community in a review called Verification, Validation, and Accreditation (VV&A). Verification is a technical requirement to make sure that the model has no conceptual or mathematical "bugs." Validation is "the process of determining the degree to which a model or simulation and its associated data are an accurate representation of the real world from the perspective of the intended uses of the model" (*8*). Although verification may be hard and tedious, similar to verifying a long and complex computer program, it is nonetheless a well-understood task. Validation is much more problematic for AC models because the "real world" is often inaccessible. Reliable combat data seldom exist in the quantity and variety needed for validation, and it is often not clear how to compare the model assumptions and results to the elusive "reality" (*9*). Current practice of validating AC models ranges from rigorous statistical methods for small-scale physical models (e.g., firing and detection models) and some guerrilla models that are supported with sufficient field data to reviews by subject-matter experts for large-scale campaign–level models (*10*). Accreditation is defined as an "official certification that a model or simulation and its associated data are acceptable for use for a specific purpose" (*10*).

Field exercises are designed for training and are as close as one can get to real combat. These exercises include real weapons and take place in open terrain. The only (significant!) deviation from reality is the absence of a real-life enemy. With proper control and refereeing, these models are effective for training military forces in tactics and command and control of military formations. War games are similar models that represent a higher level of abstraction. First, these models are executed on a simulated battlefield (such as a sand table, map, or computer screen) without any real weapons. Second, only selected key players take part in these models. The uncertain behavior of the enemy and

the capricious effects of nature are determined by human umpires (for example, by throwing dice) or randomly generated by computers. War games may be completely manual (Fig. 1A) or be played with computerized intervention, to generate a field of combat and/or link participants (Fig. 1B). Monte Carlo or agent-based simulations are fully computerized, in which decisions, usually made by human actors, are made by some preprogrammed decision rules. The applications of Monte Carlo and agent-based simulations range from analyzing tactics and the effectiveness of weapon systems in various scenarios (*11*) to studying geopolitical competitions among states and the effect of conquest (*12*).

Although popular and sometimes useful, these simulations have a major drawback that is manifested in the gap between decision-making in a real armed conflict and decisions made in a simulation. The risk of severe damage, injury, and loss of life may dictate a different attitude toward risk in real-life combat than in simulations where, say, the loss of an aircraft carrier is manifested by simply removing an icon from the computer screen (*4*).

Unlike simulations, which can deal with a large number of parameters, analytical models capture only selected key aspects of armed conflicts. They may be deterministic [such as sets of differential equations (*13*)] or stochastic [such as Markov chains (*14*)]. Lanchester proposed a family of ordinary differential equations describing the dynamics of force-on-force engagements (*13*). These equations have been used extensively in the past 60 years in warfare models developed by the U.S. Army (*11*). Stochastic versions of Lanchester models address the inherent randomness associated with combat (*15*).

The Guerrilla Warfare model (*16*) is a variation of a Lanchester model in which one side—the guerrilla force that hides or blends into the population—engages a regular force that is fully exposed. The guerrilla force uses aimed fire (like the Taliban attacks on NATO forces in Afghanistan), while the regular force has to search for the guerrilla force. This model has given insights regarding the tradeoff between fire (attrition) and intelligence (*17*). Although providing a compelling physical description of military attrition, Lanchester models have been criticized in the literature mostly because of the difficulty of empirically validating them (*18*, *19*). There have been several attempts to fit the Lanchester models to WWII battle data (*20*–*22*), with mixed results. Lanchester equations were described as red herrings in a 1987 National Academies study (*23*); however, they are still useful as a description of attrition processes. For example, the Salvo model, proposed by Hughes (*24*), is an adaptation of the Lanchester models to modern naval warfare, which is characterized by exchanges of missile salvos.

Optimization models generate decisions that are optimal in some sense and include resource allocation to tactical defense planning (*25*), sensor deployment (*26*), and engagement tactics (*27*). A weapon assignment model (*28*) has been implemented in the U.S. Navy for real-time pairing of Tomahawk missiles with designated missions. Optimization techniques include linear and integer programming (such as defense resource allocation and weapon assignment), stochastic programming (such as sensor deployment), and dynamic programming (such as engagement tactics). Attacker-defender models, where decisions are sequential—a defender acts first, an attacker observes these actions and responds accordingly—have been treated as mathematical programming (optimization) models by Brown and others (*29*). Such models encompass a wide range of combat situations such as missile defense (*30*) and protecting national infrastructure (*29*). The basic idea in these models is that the defender acts in such a way as to minimize the effect of the attacker (or maximize its own surviving assets), whereas the attacker acts optimally in opposition. The missile defense model (*30*) has been used in recommending specific tactics to defend the United States and its allies, assess the value of improved technologies, and critically evaluate alternate defense investments.

Game theory models are, in a sense, the mathematical manifestation of war games. Instead of real (human or computerized) players who run the game, game theory prescribes policies that are "optimal" and "stable" with respect to each player, without the need to actually play the game. Game theory was originally intended to mainly explain economic behavior (*31*, *32*). The Colonel Blotto game (*33*), originated in 1921 by the French mathematician Borel, is a zero-sum game that addresses a typical dilemma of each of two adversaries: how to deploy forces among several battlefields so as to maximize their overall force effectiveness. After WWII, game theory became an important tool for analyzing strategic defense issues, in particular during the Cold War (*34*).

Current and Future AC Modeling

Armed conflicts before and after WWII and the Cold War were typically wide-ranging and involved heavily armed state actors also possessing strategic weapons such as nuclear bombs. This motivated the development of large-scale force-on-force and ballistic missile defense models. Although these types of conflicts may still be feasible and require appropriate modeling, current armed conflicts are of smaller scale and also involve nonstate actors such as rebels, guerrillas, insurgents, and terrorists. These conflicts are sometimes called asymmetric or irregular war, which has been defined by the U.S. Department of Defense as "a violent struggle among state and nonstate actors for legitimacy and influence over the relevant populations" (*35*).

Terrorism is one very prevalent aspect of armed conflicts today that has triggered several studies that range from modeling individual attacks to analyzing terror-events data to developing game-theoretic models. One of the most common forms of terror attacks is person-borne suicide bombs. Studying the effect of such attacks and analyzing response policies are of great importance to local authorities and first responders. Using some elementary geometric and probabilistic considerations, and accounting for the effect of crowd blocking, (*36*) shows that the expected number of casualties in a single suicide bomber incident is bounded by around 32—an outcome that has been observed in many such events in Israel (*36*) and Iraq (*37*). Mitigation strategies are studied in (*38*). Johnson and others (*39*) examine the patterns of violence in insurgencies and terror events and identify a common pattern regarding the size distribution of such events and their timing. Their dynamic model explaining this pattern is based on the notion of coalescence and fragmentation of insurgents or terror organizations, thus producing an ecology of groups. A more recent paper (*40*) reveals a dynamical pattern of fatal terrorist attacks. This pattern, which is manifested in a power law, identifies possible escalation scenarios of such attacks. The authors establish a new metric for understanding the momentum of these attacks and the effectiveness of counterterrorist actions—a metric that appears to be stable across multiple conflicts and at different scales.

Addressing a similar problem, Kaplan (*41*) applies queueing theory and Markov processes to predict the number of undetected terror plots and estimate the rate at which such threats can be interdicted. Game theoretic models have addressed several counterterror situations (*42*). The Israeli-Hezbollah war in 2006 motivated an optimization model for operations against entrenched guerrillas (*43*). The problem of attacks by improvised explosive devices on coalition forces in Iraq led to a model for allocating clearing devices on a network of roads in the presence of a strategic adversary (*44*). Attacker-defender mathematical programming models are well suited for homeland security problems, because defensive preparations for critical infrastructure (such as the electric grid) involve large investments, visible to any intelligent attacker who would adjust his tactics accordingly (*29*).

The two main challenges to modeling counterterror operations are detection and protection. Effective intelligence collection and analysis are crucial for timely detection and interception of terror plots. Mathematical models such as Bayesian updating, Markov random fields, exploration-exploitation schemes, and probability graphical models can improve the efficiency of intelligence collection. Advances in data mining help "connect the dots," and the application of specially tailored techniques will result in better allocation of human and equipment resources and improve the work processes during the analysis phase (*45*). Modeling the vulnerabilities of a system and the effect of mitigation actions against a terror attack by a strategic adversary is essential for developing effective

protection schemes against such attacks. In particular, mathematical models should be used to provide policy recommendations in homeland security on topics such as building redundancies in national infrastructures (*29*), border screening, and biological attacks (*46*).

In one-on-one situations, the nonstate actors in an insurgency are no match for the military might of state-controlled forces, who are significantly larger and better equipped and trained. To avoid eradication, nonstate actors must hide, making themselves difficult to detect and target by state forces. This elusiveness is attained by blending in with the civilian population and using relatively simple, yet lethal, weapons such as small arms, improvised explosive devices, and even biological agents that do not require persistent exposure.

Hiding places, shelters, information, logistical support, and recruits are provided to the nonstate actors by the civilian population, either willingly or by coercion. The civilian population is also a source of information (intelligence) to the state forces, a consumer of social and economic resources, and a target of terrorist attacks. All of these characteristics make civilians a key component in irregular warfare modeling, which is absent in legacy AC models.

Following Deitchman's classic guerrilla model (*16*), more-recent work has focused on the dynamics of insurgencies, as depicted in Fig. 2. Kress and Szechtman (*47*) developed an attrition-reinforcement model representing recent wars in Iraq and Afghanistan. The model incorporates dynamic relationships among intelligence gathered, collateral casualties in the population (which can turn public opinion against the perpetrators), recruitment to the insurgency, and reinforcement to government forces. It demonstrated that, under some reasonable assumptions, an insurgency cannot be eradicated by force alone; at best, it can be contained in a stalemate, which could only be broken by nonviolent circumstances. Although reliable statistical data for validating the model is lacking, the model is consistent with available anecdotal information (*47*). Using the limited data available on the insurgencies in Malaya and Iraq, Johnson and Madin (*48*) exploit a simple population-growth logistic model to compare the dynamics of an insurgent population in these two insurgencies. They identify conditions for failure and success of an insurgency in terms of attrition and recruitment rates and the population carrying capacity (a term describing the relationship between population density and resource dependence). Berman *et al.* (*49*) model the economics of counterinsurgency, using the Empirical Studies of Conflict Project database (*50*), as a three-way contest between violent insurgents, a government seeking to minimize violence, and civilians deciding whether to share information about insurgents. The results of the model underscore the effectiveness of service provisions by the government to the population as a violence-reducing factor.

Armed revolts, such as the recent events in Libya, Yemen, and Syria, represent a somewhat different facet of irregular warfare, in which civilian demonstrations and social unrest turn into an armed conflict. A model describing this situation (*51*) suggests that, unlike classical force-on-force models, the outcome of a revolt is independent of the initial force sizes; it only depends on the fraction of the population supporting each side and on the combat effectiveness of the government forces and the rebels. This model specifies conditions for a stalemate and underscores the critical effect of foreign intervention as a game-changer, as was demonstrated in Libya.

The aforementioned papers and a few computerized simulations developed for the Department of Defense, such as COIN 1.0 and COIN 2.0 (*52*), are only initial attempts at modeling irregular warfare. COIN 2.0 is a computerized model, which includes only limited representation of violence: just casualties from improvised explosive devices and direct fire. It focuses on the social landscape in a counterinsurgency situation and models the interrelations among the coalition forces, foreign fighters, and the civilian population.

Insurgencies are combat situations in which regime forces and insurgents fight each other using a variety of weapons, such as improvised explosive devices, suicide bombs, direct fire, drone-borne missiles, and artillery. Attrition models are necessary for modeling these situations, but they are not sufficient. The big challenge is to combine attrition models; political, social, and behavioral sciences; and economic theory into a unified model.

Social and behavioral components in such a model should include social networks (*53*), which model the underlying connectivity in the population of interest, its dynamics, and its impact on the actions of state and nonstate actors (e.g., the role of Facebook in the Arab Spring that facilitated the revolution). Social network models can be used for identifying key individuals whose absence will destabilize an adversary network (*54*) in a manner analogous to the way in which advertisers target likely buyers. Models that describe the way the topology of these self-organizing networks evolves over time (*55*), and the effect of the regime's and insurgents' actions on these dynamics, are of particular interest because they can explain changes in popular behavior. Social models should also represent the spread of ideas (*56*) that affect the pace at which people change their attitudes and capture the effect of private and public preferences (*57*) that explain differences between the hidden attitude of an individual toward one side in the conflict and his exposed behavior (e.g., the effect of coercion). The effect of the mass media on public opinion, and its impact on what governments do (such as violence against civilians or foreign intervention), is also an important factor to be modeled. Economic theory models should capture the role of incentives in shaping popular behavior and buying out opponents [e.g., principal-agent models (*58*)]. Another important issue to be addressed is corruption (*59*), which is relevant for rebellious situations that stem from corrupted regimes. There is a relatively large body of important research on insurgencies

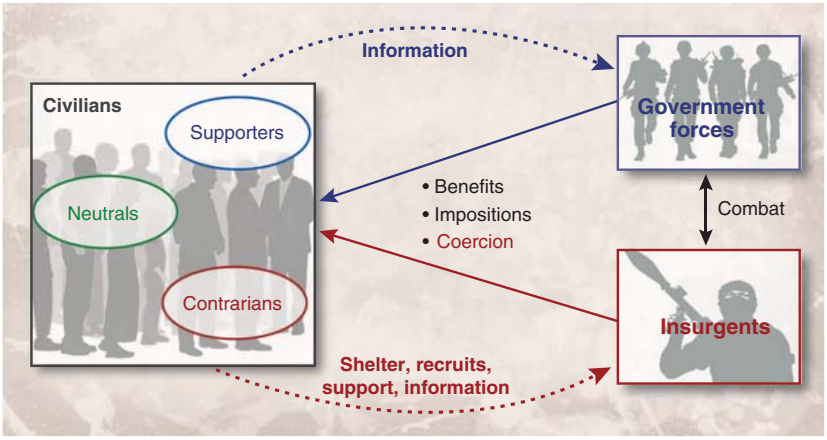


Fig. 2. Insurgency dynamics. There are three actors: the government (state) forces; the insurgency (nonstate entity); and the general population, which is divided into three sectors: supporters of the government, supporters of the insurgency (called contrarians), and neutrals. The government and the insurgents are engaged in direct combat, but both draw information from their supporters in the population. The insurgents also get other support. The size of each support group is affected by (i) social and economic incentives (such as health care and infrastructure) provided to the population (or sector(s) thereof) by each side; (ii) impositions (such as Sharia law) and coercive actions that aim to intimidate and affect the flow of information; and (iii) collateral damage caused by combat actions (such as civilians caught in crossfire) and misdirected coercion. Although in some cases the government may also execute coercive actions against civilians, in this model we assume that it does not.

in the political science literature that can be incorporated in future AC models. In particular, empirical studies of certain insurgencies shed light on the feedback effect between a regime's violence and the level of insurgency. For example, studies of the insurgency in Chechnya (*60*) and the Vietnam War (*61*) provide important insights for modeling the insurgency dynamics. In addition, data assembled in the Correlates of War project (*62*) can be used for validating future AC models, at least at the macroscopic level.

What will be needed in the future? Certainly advances in defense technology, which affect both regular and irregular warfare, will need models to assess their impact and optimize the employment of the resulting weapons and equipment. In particular, networks of sensors, which support the operation of unmanned systems, will require data fusion and machine-learning models that will facilitate an effective use of these two advanced technologies. Also, to better model and understand future armed conflicts, information technology should be implemented for systematically collecting data about the combat environment, actions, and outcomes during such events.

The emphasis in this article has been on quantitative models, but answering some broader questions may require more qualitative analysis of behavioral factors and social forces. Why does someone become a terrorist, and how can that process be stopped? How long does it take for a population to forget a multigenerational history of conflict, as in Northern Ireland or the Middle East? Insights from a range of fields, some described in this issue of *Science*, and further understanding of the complexity of human interactions during armed conflicts will need to be part of our arsenal in confronting these questions in the future.

References and Notes

1. V. S. Subrahmanian, J. Dickerson, *Science* **326**, 1201 (2009).
2. R. M. Citino, *The German Way of War: From the Thirty Years War* (Univ. of Texas Press, Austin, TX, 2005), p. 150.
3. G. F. Kimball, P. M. Morse, *Methods of Operations Research* (Wiley, New York, 1951).
4. A. Washburn, M. Kress, *Combat Modeling* (Springer, New York, 2009).
5. A. Ilachinski, *Artificial War: Multiagent-Based Simulation of Combat* (World Scientific, Hackensack, NJ, 2004).
6. EADSIM, Brown Engineering, Teledyne Technologies; accessed 15 December 2011 (www.eadsim.com/overview.asp).
7. National Research Council of the National Academies, *Defense Modeling, Simulation and Analysis* (National Academies Press, Washington, DC, 2006).
8. www.dtic.mil/whs/directives/corres/pdf/500061p.pdf
9. C. J. Thomas, in *Military Models for Decisions*, W. P. Hughes Jr., Ed. [Military Operations Research Society (MORS), Alexandria, VA, 1997], pp. 333–349.
10. http://armypubs.army.mil/epubs/pdf/P5_11.PDF.
11. S. Bonder, *Oper. Res.* **50**, 25 (2002).
12. N. B. Weidmann, L.-E. Cederman, *Soc. Sci. Comput. Rev.* **26**, 510 (2008).
13. F. W. Lanchester, *Aircraft in Warfare: The Dawn of the Fourth Arm* (Constable, London, 1916).
14. C. Barfoot, *Oper. Res.* **22**, 318 (1974).
15. M. Kress, I. Talmor, *J. Oper. Res. Soc.* **50**, 733 (1999).
16. S. Deitchman, *Oper. Res.* **10**, 818 (1962).
17. T. S. Schreiber, *Oper. Res.* **12**, 507 (1964).
18. J. W. R. Leppingwell, *Int. Secur.* **12**, 89 (1987).
19. C. A. Fowler, *Asymmetric Warfare: A Primer* (IEEE Spectrum, http://spectrum.ieee.org/aerospace/aviation/asymmetric-warfare-a-primer, March 2006).
20. N. MacKay, C. Price, *History (London)* **96**, 304 (2011).
21. R. D. Fricker, *Nav. Res. Log.* **45**, 1 (1998).
22. T. W. Lucas, T. Turkes, *Nav. Res. Log.* **51**, 95 (2004).
23. Commission on Physical Sciences, Mathematics, and Applications, the National Academies, *Technology for the United States Navy and Marine Corps, 2000-2035: Becoming a 21st-Century Force: Volume 9: Modeling and Simulation* (National Academies Press, Washington, DC, 1997); www.nap.edu/openbook.php?record_id=5869&page=227.
24. W. P. Hughes Jr., in *Warfare Modeling,* J. Bracken, M. Kress, R. E. Rosenthal, Eds. (MORS and Wiley, New York, 1996), pp. 121–143.
25. I. Ravid, *Oper. Res.* **37**, 700 (1989).
26. M. Kress, J. Royset, *Mil. Oper. Res.* **13**, 23 (2008).
27. D. P. Gaver, P. A. Jacobs, G. Samorodnitsky, K. D. Glazebrook, *Nav. Res. Log.* **53**, 588 (2006).
28. A. M. Newman *et al.*, *Nav. Res. Log.* **58**, 281 (2011).
29. G. G. Brown, W. M. Carlyle, J. Salmeron, K. Wood, in *Tutorials in Operations Research: Emerging Theory, Methods, and Applications*, H. Smith, J. Greenberg, Eds. (INFORMS, Hanover, MD, 2005), pp. 102–123.
30. G. G. Brown, M. Carlyle, D. Diehl, J. Kline, K. Wood, *Oper. Res.* **53**, 745 (2005).
31. J. v. Neumann, *Math. Ann.* **100**, 295 (1928).
32. J. von Neumann, O. Morgenstern, *Theory of Games and Economic Behavior* (Princeton Univ. Press, Princeton, NJ, 1944).
33. R. Bellman, *SIAM Rev.* **11**, 66 (1969).
34. M. Dresher, *Games of Strategy: Theory and Application* (RAND Corporation, Santa Monica, CA, 2007).
35. United States Department of Defense, *Irregular Warfare: Countering Irregular Threats. Joint Operating Concept Version 2.0 17 May 2010* (Irregular Warfare Joint Operations Command, 2010), p. 9.
36. M. Kress, *Nav. Res. Log.* **52**, 22 (2005).
37. www.iraqbodycount.org/database
38. E. H. Kaplan, M. Kress, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 10399 (2005).
39. J. C. Bohorquez, S. Gourley, A. R. Dixon, M. Spagat, N. F. Johnson, *Nature* **462**, 911 (2009).
40. N. Johnson *et al.*, *Science* **333**, 81 (2011).
41. E. H. Kaplan, *Oper. Res.* **58**, 773 (2010).
42. T. Sandler, K. Siqueira, *Simul. Gaming* **40**, 164 (2009).
43. E. H. Kaplan, M. Kress, R. Szechtman, *Oper. Res.* **58**, 329 (2010).
44. A. Washburn, P. L. Ewing, *Nav. Res. Log.* **58**, 180 (2011).
45. E. H. Kaplan, in *Intelligence Analysis: Behavioral and Social Scientific Foundations,* B. Fischhoff, C. Chauvin, Eds. (National Academies Press, Washington, DC, 2011), pp. 31–56.
46. L. Wein, *Oper. Res.* **57**, 801 (2009).
47. M. Kress, R. Szechtman, *Oper. Res.* **57**, 578 (2009).
48. D. D. P. Johnson, J. S. Madin, in *Natural Security: A Darwinian Approach to a Dangerous World*, R. D. Sagarin, T. Taylor, Eds. (Univ. of California Press, Berkeley, CA, 2008), pp. 159–185.
49. E. Berman, J. N. Shapiro, J. H. Felter, *J. Polit. Econ.* **119**, 766 (2011); www.princeton.edu/~jns/publications/BSF_2011_HAM.pdf.
50. www.hoover.org/news/28932
51. M. P. Atkinson, A. Gutfraind, M. Kress, *J. Oper. Res. Soc.*, published online 21 December 2011 (10.1057/jors.2011.146).
52. Z. A. Henscheid, M. T. K. Koehler, S. K. Mulutzie, B. F. Tivnan, J. G. Turnley, *COIN 2.0 Formulation* (MITRE, Report No. 10-4822, 2010).
53. M. O. Jackson, *Social and Economic Networks* (Princeton Univ. Press, Princeton, NJ, 2008).
54. www.casos.cs.cmu.edu/publications/working_papers/Carley-NAACSOS-03.pdf
55. D. J. Watts, S. H. Strogatz, *Nature* **393**, 440 (1998).
56. L. M. A. Bettencourt, A. Cintron-Arias, D. I. Kaiser, C. Castillo-Chavez, *Physica A* **364**, 513 (2006).
57. T. Kuran, *Public Choice* **61**, 41 (1989).
58. J.-J. Laffont, D. Martimort, *The Theory of Incentives: The Principal-Agent Model* (Princeton Univ. Press, Princeton, NJ, 2002).
59. S. Rose-Ackerman, *Corruption: A Study in Political Economy* (Academic Press, New York, 1978).
60. J. Lyall, *J. Conflict Resolut.* **53**, 331 (2009).
61. M. A. Kocher, T. B. Pepinsky, S. N. Kalyvas, *Am. J. Pol. Sci.* **55**, 201 (2011).
62. http://144.206.159.178/ft/CONF/16432648/16432658.pdf

Acknowledgments: I thank G. Brown for his meticulous review of several drafts of this paper and his important suggestions. I also thank the five reviewers for their constructive and insightful remarks. Finally, thanks to J. Appleget, M. Atkinson, R. Dell, W. Hughes, J. Kline, N. Kress, T. Lucas, R. Szechtman, and A. Washburn for their valuable comments and inputs.

10.1126/science.1217724

PERSPECTIVE

CLIMATE CHANGE AND VIOLENT CONFLICT

Jürgen Scheffran,[1]* Michael Brzoska,[2] Jasmin Kominek,[1,3] P. Michael Link,[1,4] Janpeter Schilling[1,5]

Current debates over the relation between climate change and conflict originate in a lack of data, as well as the complexity of pathways connecting the two phenomena.

Since publication of the fourth assessment report of the Intergovernmental Panel on Climate Change (IPCC), the debate on the security implications of climate change has intensified. Research in this area has made progress but remains controversial [for recent reviews, see (*1*–*4*)]. Although some quantitative empirical studies support a link between climate change and violent conflict, others find no connection or only weak evidence.

A major challenge for all studies is to find adequate data. Instead of using data on the long-term average and variability of temperature, precipitation, and other climatic variables that would clearly fall under the IPCC definition of climate change (*5*), many studies have used proxies, such as short-term data on weather and extreme weather events, or on natural phenomena of climate variability like the El Niño Southern Oscillation (*6*).

It is important to distinguish between the types of conflict used in various data sets. The widely used Armed Conflict Dataset of the Uppsala Conflict Data Program and the Peace Research Institute Oslo (UCDP-PRIO), for instance, sets a minimum of 25 battle-related deaths per year and involvement of at least one state government to be considered as armed conflict (*7*). This excludes other forms of violent or nonviolent behavior that may be affected by climate change such as protests, riots, or livestock theft, let alone conflict as a positional difference over interests, values, or goals. These distinctions are relevant as, in recent decades, climate variability may have been more associated with low-level violence and internal civil war—which fall below the UCDP-PRIO definition cutoff—than with armed conflict or war between countries.

Long-term historical studies tend to find a coincidence between climate variability and armed conflict, in line with some narratives about the evolution and collapse of civilizations [e.g., (*8*)]. For instance, Zhang and others (*9*) combine a set of variables for the time period 1500–1800 to identify climate change as a major driver of large-scale human crises in the Northern Hemisphere. Tol and Wagner (*10*) cautiously conclude that, in preindustrial Europe, cooler periods were more likely to be related to periods of violence than warmer phases. Similar findings have been presented for eastern China (*11*).

However, the results have been less conclusive for recent periods. For instance, in one study, a significant correlation between temperature and civil war in Africa between 1981 and 2002 is used to project a substantial climate-induced increase in the incidence of civil war in Africa until 2030 (*12*). Yet, this result is not robust for an extended time period and alternative definitions of violent conflict (*13*).

Food insecurity has been found to contribute to violence, as exemplified by recent "food riots" (*14*, *15*), but there is little empirical evidence that climate variability is an important driver of violent land-use conflicts—e.g., in the Sahel (*16*). In Kenya, changing rainfall patterns have the potential to increase resource scarcity as a driver of pastoral conflict (*17*). However, more conflict in the form of violent livestock theft is reported during the rainy season than during drought (*18*).

Similarly, conflicts over shared river systems have been associated with low-level violence, yet full-scale wars are unlikely [e.g., (*19*, *20*)]. Instead, an increase in international water agreements has been observed (*21*).

Finally, some studies suggest that natural disasters related to extreme weather conditions substantially increase the risk of intrastate conflict (*22*). In contrast, Bergholt and Lujala (*23*) find no increased likelihood of armed civil conflict due to weather-related disasters, and Slettebak (*24*) observes that, in crisis, cooperation frequently prevails.

New research is on the way as new databases on nonstate conflicts, low-level violence, social instability events, and geo-referenced spatio-temporal patterns become available (*25*–*27*) (table S1). In addition to data needs, it is important to account for complexities in the relation between climate change and conflict. There are multiple pathways and feedbacks between the climate system, natural resources, human security, and societal stability (Fig. 1).

Since the 1990s, there has been an extensive scientific debate on how the scarcity of natural resources affects violence and armed conflict (*29*, *30*). More recently, conflict studies pay attention to the vulnerability of natural and social systems to climate impacts (*31*). Vulnerability can be broken down into three factors: (i) exposure to climate change, (ii) sensitivity to climate change, and (iii) adaptive capacity (*32*). The last two can be affected by conflict. Many of the world's poorest people are exposed to various risks to life, health, and well-being. If climate change adds to these risks, it can increase humanitarian crises and aggravate existing conflicts without directly causing them.

The question is whether human development, resilience, and adaptive capacity can compensate for increasing exposure and sensitivity to climate change. In previous decades, humanitarian aid, development assistance, and wealth per capita have increased (*33*), which has contributed to a reduction of global poverty as a possible driver of conflict. International efforts to prevent and manage conflicts have also been strengthened, and the number of armed conflicts has declined since the end of the Cold War (*34*). In recent years, however, this trend slowed down or is being reversed. While the number of democratic states has grown over the past half-century, the number of fragile states with weak institutions has also increased (*35*).

If the debate on the securitization of climate change provokes military responses and other extraordinary measures, this could reinforce the likelihood of violent conflict. Main aspects of security concern include interventions in fragile

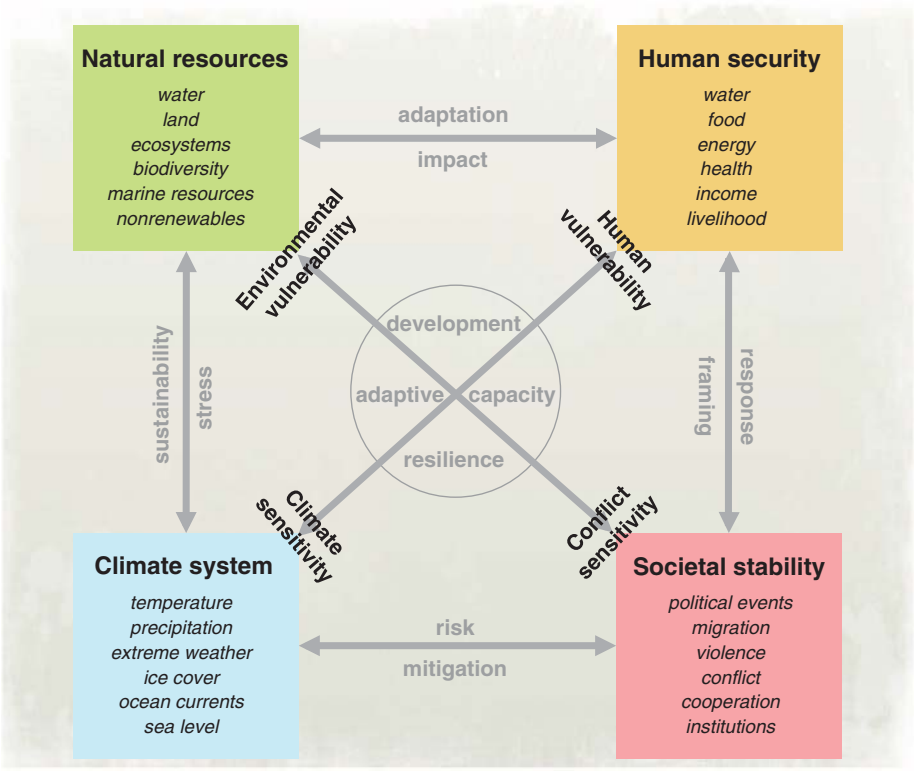


Fig. 1. Analytical framework of linkages between the climate system, natural resources, human security, and societal stability [based on (*28*)].

[1]Research Group Climate Change and Security, Institute of Geography and KlimaCampus, University of Hamburg, Grindelberg 5, D-20144 Hamburg, Germany. [2]Institute for Peace Research and Security Policy and KlimaCampus, University of Hamburg, Beim Schlump 83, D-20144 Hamburg, Germany. [3]Institute of Sociology, University of Hamburg, Allende-Platz 1, D-20146 Hamburg, Germany. [4]Research Unit Sustainability and Global Change, Center for Earth System Research and Sustainability, University of Hamburg, Bundesstrasse 55, D-20146 Hamburg, Germany. [5]School of Integrated Climate System Sciences, KlimaCampus, University of Hamburg, Grindelberg 5, D-20144 Hamburg, Germany.

*To whom correspondence should be addressed. E-mail: juergen.scheffran@zmaw.de

Table 1. List of core research questions structured by the relations between causes and effects of the human-environment interaction.

	Effect			
Cause	**Climate change**	**Natural resources**	**Human security**	**Societal stability**
Climate change	Which climate feedbacks enhance or dampen the speed of climate change? Where are thresholds and tipping points?	How are water, land, and biodiversity affected by climate change, e.g., by drought, soil erosion, or flooding?	How do extreme weather and climate variability affect human livelihoods, health, income, and assets?	How does extreme weather affect social conflicts? How can research scenarios of impacts inform politics?
Natural resources	How do losses of natural resources affect climate change, e.g., through deforestation, ocean uptake, or desertification?	Are there relevant natural adaptation or substitution processes for the loss of natural resources?	How does resource availability affect human security? How to increase resilience and adaptive capacity?	Is conflict triggered by resource abundance or degradation? Does societal stability depend on natural resources?
Human security	Under which conditions do gains or losses of human security drive climate change and mitigation?	How does human (in-)security affect the use of natural resources? Does a decline in production reduce resource inputs?	Do elements of human (in-)security reinforce each other? Will security risks spread to neighbor regions?	Does human insecurity drive cooperation or conflict? Will human responses lead to social transformation?
Societal stability	How do social unrest and violent conflict affect carbon emissions? Will societal stability lead to climate mitigation?	How does societal stability affect resource exploitation? Can cooperation protect resource stocks?	How do conflict, societal instability, and cooperation affect human security and vulnerability?	Under which conditions do societies (de-)stabilize themselves or solve conflicts? What is the role of institutions?

states, the securing of borders (e.g., against disaster refugees), and access to resources (e.g., in the Mediterranean or Arctic region) [see (*36*)]. Other responses to climate change may also become causes of conflict, including bioenergy (as producers compete for land and food-related resources), nuclear power (which can lead to nuclear weapons proliferation), or geoengineering (through disagreements between states). Thus, there is a need for conflict-sensitive mitigation and adaptation strategies that contain conflict and contribute to cooperation via effective institutional frameworks, conflict management, and governance mechanisms.

Research Challenges

The balance between political and social factors and climate change could shift when the global temperature reaches levels that have been unprecedented in human history. There is reason to believe that such a change might overwhelm adaptive capacities and response mechanisms of both social and natural systems and thus lead to "tipping points" toward societal instability and an increased likelihood of violent conflict (*37*).

Although some fundamental issues have been raised in previous research, numerous interdisciplinary questions still need to be investigated to understand the feedback loops involved (Table 1). Models of the various linkages can build on a rich set of tools from complexity science, multiagent systems, social-network analysis, and conflict assessment to extend previous data and experiences into future scenarios that cover different social, economic, and political contexts (*28*). Research across scientific disciplines will be needed to identify opportunities and coherent strategies to address societal challenges related to climate change.

References and Notes

1. N. P. Gleditsch, *J. Peace Res.* **49**, 3 (2012).
2. J. Scheffran, M. Brzoska, H. G. Brauch, P. M. Link, J. Schilling, Eds., *Climate Change, Human Security and Violent Conflict: Challenges for Societal Stability* (Springer, Berlin, 2012).
3. T. Bernauer, T. Böhmelt, V. Koubi, *Environ. Res. Lett.* **7**, 015601 (2012).
4. H. Buhaug, O. M. Theisen, in (*2*), pp. 43–55.
5. IPCC, *Climate Change 2007: The Physical Science Basis* (Cambridge Univ. Press, Cambridge, 2007).
6. S. M. Hsiang, K. C. Meng, M. A. Cane, *Nature* **476**, 438 (2011).
7. UCDP-PRIO Armed Conflict Dataset Codebook Version 4-2011 (2011); http://www.pcr.uu.se/digitalAssets/63/63324_Codebook_UCDP_PRIO_Armed_Conflict_Dataset_v4_2011.pdf.
8. R. Kuper, S. Kröpelin, *Science* **313**, 803 (2006).
9. D. D. Zhang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 17296 (2011).
10. R. Tol, S. Wagner, *Clim. Change* **99**, 65 (2010).
11. D. D. Zhang, J. Zhang, H. F. Lee, Y. He, *Hum. Ecol.* **35**, 403 (2007).
12. M. B. Burke, E. Miguel, S. Satyanath, J. A. Dykema, D. B. Lobell, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 20670 (2009).
13. H. Buhaug, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 16477 (2010).
14. E. Messer, *NAPA Bull.* **32**, 12 (2009).
15. S. Johnstone, J. Mazo, *Survival (London)* **53**, 11 (2011).
16. T. A. Benjaminsen, K. Alinon, H. Buhaug, J. T. Buseth, *J. Peace Res.* **49**, 97 (2012).
17. F. E. O. Opiyo, O. V. Wasonga, J. Schilling, S. Mureithi, *Afric. J. Environ. Sci. Technol.* (2012).
18. W. R. Adano, T. Dietz, K. Witsenburg, F. Zaal, *J. Peace Res.* **49**, 65 (2012).
19. M. Brochmann, P. R. Hensel, *Int. Negot.* **14**, 393 (2009).
20. T. Bernauer, T. Siegfried, *J. Peace Res.* **49**, 227 (2012).
21. L. De Stefano *et al.*, *J. Peace Res.* **49**, 193 (2012).
22. P. Nel, M. Righarts, *Int. Stud. Q.* **52**, 159 (2008).
23. D. Bergholt, P. i. Lujala, *J. Peace Res.* **49**, 147 (2012).
24. R. T. Slettebak, *J. Peace Res.* **49**, 163 (2012).
25. C. Raleigh, D. Kniveton, *J. Peace Res.* **49**, 51 (2012).
26. P. F. Nardulli, K. H. Leetaru, in (*2*), pp. 57–76.
27. J. W. Busby, T. G. Smith, K. L. White, S. M. Strange, in (*2*), pp. 463–512.
28. J. Scheffran, P. M. Link, J. Schilling, in (*2*), pp. 91–131.
29. T. Homer-Dixon, *Int. Secur.* **19**, 5 (1994).
30. G. Bächler, in *Environmental Change and Security: A European Perspective*, A. Carius, K. M. Lietzmann, Eds. (Springer, Berlin, 1999), pp. 107–130.
31. J. Scheffran, A. Battaglini, *Reg. Environ. Change* **11** (suppl. 1), 27 (2011).
32. IPCC, *Climate Change 2007: Climate Change Impacts, Adaptation and Vulnerability* (Cambridge Univ. Press, Cambridge, 2007).
33. Organization for Economic Cooperation and Development, Analyses of aid—How much aid is delivered, where, and for what purpose (2012); www.oecd.org/document/48/0,3746,en_2649_34447_42396656_1_1_1_1,00.html.
34. L. Themnér, P. Wallensteen, *J. Peace Res.* **48**, 525 (2011).
35. M. G. Marshall, K. Jaggers, Polity IV Project: Political Regime Characteristics and Transitions, 1800–2010 (2010); www.systemicpeace.org/inscr/inscr.htm.
36. M. Brzoska, in (*2*), pp. 165–184.
37. WBGU, *World in Transition—Climate Change as a Security Risk* (Earthscan, London, 2008).

Acknowledgments: Research for this study was funded in part by the German Science Foundation (DFG) through the Cluster of Excellence "CliSAP" (EXC177).

Supplementary Materials

www.sciencemag.org/cgi/content/full/336/6083/869/DC1
Table S1
References

10.1126/science.1221339

HUMANCONFLICT
http://scim.ag/hum_conflict

PERSPECTIVE

ARE WE WINNING THE WAR AGAINST POSTTRAUMATIC STRESS DISORDER?

Richard J. McNally

The most methodologically rigorous epidemiological study on American military personnel deployed to Iraq and Afghanistan found that 4.3% of troops developed posttraumatic stress disorder (PTSD). Among deployed combatants, 7.6% developed PTSD, whereas 1.4% of deployed noncombatants did so. The U.S. Department of Veterans Affairs has launched a program ensuring that all veterans with PTSD will receive evidence-based cognitive-behavioral therapy, and the Army has developed Battlemind postdeployment early interventions that reduce risk for the disorder.

The outbreak of war in Afghanistan and Iraq prompted dire predictions about its likely psychiatric consequences. The chief of readjustment counseling services at the U.S. Department of Veterans Affairs (VA) conjectured that as many as 30% of troops deployed to Iraq might develop posttraumatic stress disorder (PTSD) (*1*), a syndrome that can emerge after exposure to horrific, life-threatening events, such as combat, natural disasters, and rape. PTSD sufferers do not merely remember their trauma; they reexperience it as vivid sensory recollections (flashbacks), nightmares, and intrusive thoughts. They feel numb and emotionally disconnected from loved ones, yet also tense, irritable, and hypervigilant as if danger were forever present.

Psychiatry ratified the PTSD diagnosis in 1980, chiefly in response to the belated recognition of its symptoms in Vietnam veterans whose problems had long been inadequately understood and treated. Indeed, the most rigorous epidemiological study ever done on Vietnam veterans had reported that 30.9% of men who served in this war developed PTSD (*2*), furnishing a basis for early predictions about PTSD among Iraq veterans. Keen to avoid the mistakes of the Vietnam era, American, British, and Dutch authorities launched epidemiological surveys assessing the mental health of troops returning from Iraq and Afghanistan, aiming to ascertain the prevalence of PTSD and enable its early detection and treatment.

The Epidemiology of PTSD

A decade later, the data are in, and the implications are surprisingly optimistic. The wars have certainly caused PTSD, but at rates far lower than many had expected. The most methodologically sound surveys have assessed large numbers of military personnel (or veterans) randomly sampled from the overall population of American and British troops who have served in Iraq and Afghanistan. These studies show that the proportion of troops that has developed PTSD ranges from 2.1 to 13.8%. (*3*)

The most rigorous study on American troops is the U.S. Millennium Cohort study, a population-based, longitudinal investigation of active duty and Reserve/National Guard personnel (*4*). It involves random samples representative of the subpopulations of deployed combatants, deployed noncombatants, and nondeployed noncombatants. Hence, it avoids the biases associated with convenience samples or from samples drawn from those seeking treatment. Moreover, the subjects were free of PTSD at baseline, providing estimates of PTSD attributable to military trauma alone and not to trauma occurring before a soldier's enlistment in the service. Assessing 47,837 members of the Armed Forces, the researchers found that 4.3% of personnel deployed to Afghanistan or Iraq developed PTSD. Among deployed personnel, 7.6% of those reporting combat exposure developed the disorder, whereas 1.4% did so among those not experiencing combat. Of those who had never deployed overseas, 2.3% developed PTSD in response to stateside trauma (such as accidents on military bases). To be sure, rates of 4.3% among all deployers and 7.6% among combatants are not trivial. Yet, these figures are much lower than the predicted figure of 30% for all deployed troops, noncombatants as well as combatants (*1*, *2*).

A longitudinal study involving a subset of the Millennium Cohort provides further reason for cautious optimism. The researchers assessed PTSD symptoms before deployment to Afghanistan or Iraq and at two postdeployment follow-up assessments separated by a 3-year interval, enabling them to track the course of symptoms over time. (*5*) Among single- and multiple-deployers, respectively, 6.7 and 4.5% were healthy at baseline but were symptomatic at both follow-up assessments, which is suggestive of chronic PTSD directly attributable to war. Yet, most soldiers deploying either once (83.1%) or multiple times (84.9%) were resilient, exhibiting a PTSD-free healthy trajectory across all three assessment points.

One limitation of these surveys is their reliance on questionnaire measures of PTSD. Although cost effective, this approach can overestimate PTSD relative to the "gold standard" of a structured clinical interview. For example, a study of 382 Dutch infantry veterans of Iraq yielded a questionnaire-based rate of PTSD of 21%, whereas structured interviews involving 339 of them revealed a rate of 4% (*6*). Some veterans apparently misunderstood certain questions, answered questions in reference to events unrelated to the war (such as receiving upsetting news from home), or experienced insufficient impairment to qualify for the diagnosis. Overestimates are less likely when questionnaire studies require that symptoms surpass stringent severity thresholds and that they produce social and occupational impairment.

The Treatment of PTSD

Unexpectedly modest rates of PTSD among recent veterans do not justify complacency about the problem. Regardless of the prevalence of PTSD, removing obstacles to prompt, efficacious treatment is essential.

One obstacle is stigma about seeking help for mental health problems, which is a common concern among active duty troops, especially for those combatants reporting the most symptoms on anonymous surveys. Encouragingly, worries about stigma diminish in military units characterized by strong cohesion and excellent leadership (*7*).

Concern about stigma may lessen after personnel separate from the service. Among American veterans of Afghanistan and Iraq seeking any form of health care from the VA, 25% receive a mental health diagnosis, and 52% of these have PTSD (*8*). These data indicate that not all mental health problems constitute PTSD; others include depression, alcohol abuse, and difficulties readjusting to family life.

Nevertheless, the importance of PTSD has inspired a landmark VA initiative to ensure that veterans with the disorder receive either prolonged exposure (PE) or cognitive processing therapy (CPT) (*9*). Both are cognitive-behavioral therapies that have the strongest evidential support for treating the disorder, at least among civilian trauma victims. Before this initiative, less than 10% of VA clinicians specializing in the treatment of PTSD routinely used PE.

PE requires patients to recount traumatic memories repeatedly within a structured, supportive therapeutic context until distress declines. In addition to this imaginal exposure component, PE also involves gradual, systematic exposure to feared, but safe, reminders of the trauma in everyday life. CPT has patients recount their traumatic experiences repeatedly in writing. Both PE and CPT require therapists to identify and help

Department of Psychology, Harvard University, Cambridge, MA 02138, USA. E-mail: rjm@wjh.harvard.edu

patients' correct maladaptive beliefs about their symptoms and about the trauma.

To ensure widespread dissemination, the VA asked the developers of PE and CPT, Edna B. Foa and Patricia A. Resick, respectively, to oversee training workshops followed by intensive supervision of cases. The original subset of VA clinicians are now themselves training additional therapists, expanding the pool of clinicians equipped to deliver state-of-the-art, evidenced-based cognitive behavior therapy. Vital to this endeavor has been administrative support for the program, guaranteed time for therapists to deliver treatment optimally, and incentives and directives to ensure its maintenance.

Although randomized controlled trials (RCTs) are essential for confirming the efficacy of the program for recent veterans, initial data from one hospital are promising. Of the 66% of veterans of Iraq and Afghanistan who completed the PE program, 74% had posttreatment PTSD scores that fell well below the clinical cutoff for PTSD (*10*). If this study is any indication, veterans of the recent wars may have much better chances of recovery than did veterans of previous wars for whom PE was unavailable.

The Prevention of PTSD

The emotional and financial costs of chronic PTSD are substantial. Ideally, prevention would be better than is treatment only after PTSD develops. This proactive approach characterizes the Comprehensive Soldier Fitness (CSF) program, which is delivered to all members of the Army regardless of their occupational specialty (such as clerk or combat infantryman) (Fig. 1). Just as training in boot camp builds physical fitness, CSF aims to build psychological fitness, drawing on principles of positive psychology to cultivate resilience in soldiers (*11*). Important goals include facilitating personal growth and inculcating skills to reduce risk for PTSD. Moreover, the CSF includes a module to help soldiers and their families manage the stress of prolonged deployments overseas (*12*).

Unfortunately, the Army implemented the program without first conducting a RCT to test whether CSF reduces the incidence of PTSD. That is, it would have been desirable to randomly assign certain brigades to receive the program and test whether it reduces rates of PTSD below that of brigades randomly assigned to receive the Army's standard program (*13*). Without such pilot testing, it will be difficult to tell whether any beneficial outcomes are attributable to CSF. In fact, RCTs have shown that some prevention programs in the mental health field have had unintended adverse consequences (*14*).

To say that we ought to prevent PTSD implies that we can do so, and it remains unclear how malleable risk factors for PTSD really are. Because most military personnel do not develop the disorder, one might argue that we should not allocate resources to efforts to prevent PTSD in people unlikely to develop the disorder in the first place, including soldiers whose duties seldom place them in harm's way.

Hence, another option is to develop a preventive intervention, test whether it works, and then deliver it to high-risk groups. Targeting groups, not individuals, would also diminish the likelihood of stigma. Exemplifying this approach, the Army has developed postdeployment "Battlemind debriefing," which is an early intervention program for preventing psychological problems among platoons returning from combat duty (*15*). Unlike other debriefing methods that may impede recovery from trauma (*16*), Battlemind deemphasizes cathartic sharing of trauma stories within the group and focuses instead on the skills needed for a successful transition from the combat zone to home. Group facilitators remind soldiers to trust their training, emphasizing how their mastery of military skills will render them resilient, and they provide tips on coping with the three common concerns of sleep difficulties, anger control, and interpersonal withdrawal.

Fig. 1. Comprehensive Soldier Fitness logo. [Credit: U.S. Army]

The Army has also developed postdeployment "Battlemind training" for larger groups of returning combat units. In this didactic intervention, leaders reframe common difficulties, such as hypervigilance, sleep difficulties, and emotional withdrawal that occur from the transition to home as adaptive combat skills that require adjustment. For example, leaders remind soldiers to apply the emotional bonding skills that they used in their combat units to reaffirm positive family relationships. They also emphasize positive cognition and coping skills.

A RCT revealed that both postdeployment Battlemind programs produced favorable outcomes at 4-month follow-up relative to the Army's standard postdeployment stress education program (*15*). Soldiers with the most combat exposure received the most benefit, reporting fewer symptoms of PTSD and depression, less difficulty sleeping, and less concern about stigma. Hence, soldiers most at risk for PTSD received the most benefit from this preventive intervention.

A RCT testing an anglicized version of postdeployment Battlemind training among UK military personnel returning from Afghanistan reduced binge drinking at 6-month follow-up relative to the standard stress debriefing but did not affect PTSD symptoms (*17*). However, the level of baseline PTSD symptoms was substantially lower than in the American RCT. Hence, there was little room for improvement.

Conclusions

Dire predictions notwithstanding, the vast majority of troops deployed to Iraq and Afghanistan have been resilient, and the prospects for recovery for those who have developed PTSD have never been better as the VA ensures that veterans receive the best evidence-based care. Some preventive interventions show promise, too. Yet, the most hopeful development is the remarkable decline in the frequency, duration, and lethality of war, especially during the past 60 years (*18*). There are multiple likely causes of this decline (*19*), but identifying those subject to control is vital for fostering this positive trend. Indeed, steps that further the global decline in violence provide the surest route to preventing PTSD throughout the world today.

References and Notes

1. S. Dentzer, Interview with Alfonso Batres. Online News Hour, available at www.pbs.org/newshour/bb/health/jan-june04/batres_extended.html; accessed 3 May 2006 (2003).
2. R. A. Kulka *et al.*, *Trauma and the Vietnam War: Report of Findings from the National Vietnam Veterans Readjustment Study* (Brunner/Mazel, New York, 1990).
3. J. Sundin, N. T. Fear, A. Iversen, R. J. Rona, S. Wessely, *Psychol. Med.* **40**, 367 (2010).
4. T. C. Smith *et al.*; Millennium Cohort Study Team, *BMJ* **336**, 366 (2008).
5. G. A. Bonanno *et al.*, *Br. J. Psychiatry* **200**, 317 (2012).
6. I. M. Engelhard *et al.*, *Br. J. Psychiatry* **191**, 140 (2007).
7. K. M. Wright *et al.*, *Psychol. Serv.* **6**, 108 (2009).
8. K. H. Seal, D. Bertenthal, C. R. Miner, S. Sen, C. Marmar, *Arch. Intern. Med.* **167**, 476 (2007).
9. B. E. Karlin *et al.*, *J. Trauma. Stress* **23**, 663 (2010).
10. P. W. Tuerk *et al.*, *J. Anxiety Disord.* **25**, 397 (2011).
11. R. Cornum, M. D. Matthews, M. E. P. Seligman, *Am. Psychol.* **66**, 4 (2011).
12. J. M. Gottman, J. S. Gottman, C. L. Atkins, *Am. Psychol.* **66**, 52 (2011).
13. R. G. Tedeschi, R. J. McNally, *Am. Psychol.* **66**, 19 (2011).

14. G. A. Bonanno, M. Westphal, A. D. Mancini, *Annu. Rev. Clin. Psychol.* **7**, 511 (2011).
15. A. B. Adler, P. D. Bliese, D. McGurk, C. W. Hoge, C. A. Castro, *J. Consult. Clin. Psychol.* **77**, 928 (2009).
16. R. J. McNally, R. A. Bryant, A. Ehlers, *Psychol. Sci. Public Interest* **4**, 45 (2003).
17. K. Mulligan *et al.*, *J. Consult. Clin. Psychol.* (2012).
18. J. S. Goldstein, *Winning the War on War: The Decline of Armed Conflict Worldwide* (Dutton, New York, 2011).
19. S. Pinker, *The Better Angels of Our Nature: Why Violence Has Declined* (Viking, New York, 2011).

Acknowledgments: I thank N. Breslau, D. M. Wegner, and S. Wessely for helpful comments on previous drafts of this article.

10.1126/science.1222069

PERSPECTIVE

THE ANTIQUITY OF EMPATHY

Frans B. M. de Waal

The view of humans as violent war-prone apes is poorly supported by archaeological evidence and only partly supported by the behavior of our closest primate relatives, chimpanzees and bonobos. Whereas the first species is marked by xenophobia, the second is relatively peaceful and highly empathic in both behavior and brain organization. Animal empathy is best regarded as a multilayered phenomenon, built around motor mirroring and shared neural representations at basal levels, that develops into more advanced cognitive perspective-taking in large-brained species. As indicated by both observational and experimental studies on our closest relatives, empathy may be the main motivator of prosocial behavior.

After the devastations of World War II, humans were routinely depicted as "killer apes"—in contrast to the real apes, which were regarded as pacifists. Books by Konrad Lorenz, the Austrian ethologist, and Robert Ardrey, an American journalist, contributed to the idea that a hallmark of humanity is aggression. Until well into the 1980s, this remained the dominant theme of biological approaches to human behavior. This literature is now recognized as one-sided because it overlooked our species' capacity for cooperation, empathy, and prosocial behavior.

Species-typical tendencies normally come with built-in rewards. Nature has ensured that we find fulfillment in eating, sex, nursing, and socializing, all of which are necessary for survival and reproduction. If there were truly a genetic basis to our participation in lethal combat, we should willingly engage in it. Yet soldiers report a deep revulsion to killing and shoot at the enemy only under pressure (*1*). After these experiences, they often end up with substantial psychological damage. Far from being a recent phenomenon, haunting memories of combat were already known to the ancient Greeks, such as Sophocles, who described Ajax's "divine madness," now known as posttraumatic stress disorder (PTSD).

Even though evidence for individual murder goes back hundreds of thousands of years, comparable signs of warfare (such as graveyards with weapons embedded in a large number of skeletons) are lacking from before the Agricultural Revolution [about 12,000 years ago (*2*)]. This is not to imply that war was absent before then, but it does mean that the common assumption that our ancestors waged perpetual wars and knew peace only at "precarious interludes" (as Winston Churchill surmised) lacks solid archaeological backing. During most of our prehistory, we were nomadic hunter-gatherers, whose cultures are nowadays not particularly known for warfare (*3*). They do occasionally raid, ambush, and kill their neighbors (*4*), but more often trade with them, intermarry, and permit travel through their territories. Hunter-gatherers illustrate a robust potential for peace and cooperation.

Going back farther in time, we end up with *Ardipithecus ramidus*, a 4.4-million-year-old hominin that has been described as relatively peaceful, owing to its reduced canine teeth as compared to those of the chimpanzee (*Pan troglodytes*) (*5*), who can be lethally violent during territorial encounters between communities. However, the conclusion drawn from *Ardipithecus*' dentition that our ancestors were less war-prone than the apes is not rigorous unless the bonobo (*P. paniscus*), which also has relatively small canines (Fig. 1), is included. Despite being as closely related to us as chimpanzees, the behavior of bonobos fails to support traditional violence-based scenarios of human evolution. Deadly aggression among bonobos has thus far not been observed, neither in captivity nor in the wild, and xenophobia is only weakly developed. Bonobos sometimes mingle across territorial borders, where they engage in sex, grooming, and play. They are known as the "make love, not war" primates for solving dominance issues through sexual activity (*6*). Indeed, it has been suggested that these apes "may approach more closely to the common ancestor of chimpanzees and man than does any living chimpanzee" (*7*).

In addition, developments in psychology, neuroscience, behavioral economics, and animal behavior have begun to question the view, dominant until a decade ago, that animal life, and by extension human nature, is based on unmitigated competition. In primatology, the countermovement started with research into the survival value of friendships (*8*) and conflict resolution (*9*). After the discovery that chimpanzees often kiss and embrace shortly after a fight within their group, numerous studies have documented "reconciliations" in nonhuman primates. Methodologies comparing postconflict observations with baseline data to determine how species members behave in the presence versus absence of previous antagonism show that primates are generally attracted to former opponents, seeking friendly contact especially if they otherwise enjoy a mutually beneficial relationship. Relationship value appears to drive post-conflict repair (*10*). The behavioral expression of reconciliation varies, but its general effect is a rapid return to preexisting levels of tolerance and affiliation. This reunion process has been reported for macaques, gorillas, golden monkeys, capuchins, and many other primates, but also for nonprimates, such as wolves, dolphins, and hyenas. Reconciliation is a common social mechanism that would be superfluous if social life were ruled entirely by dominance and competition.

The level of cooperation among nonhuman primates tends to be underappreciated. In order to set it apart from human cooperation with nonrelatives, aid among primates is sometimes ascribed largely to kinship (*11*). This claim has not held up, however, on the basis of DNA extracted from chimpanzee feces in the wild. Males without genetic ties make up the majority of mutually supportive partnerships (*12*). The same seems to apply to bonobos. Female bonobos maintain a close social network that allows them to collectively dominate the majority of males despite the fact that females are also the migratory sex, which means that they are largely unrelated within each community (*6*). Both of our closest primate relatives are marked, therefore, by high levels of nonkin cooperation, probably explained by well-developed reciprocity.

Expressions of empathy are common in apes and resemble those of our own species. In child research, for example, a family member is typically instructed to feign distress or pain, upon which touching, stroking, and close-up eye-contact by the child is interpreted as a sign of sympathetic concern. In chimpanzees, bystanders at a fight go over to the loser and put an arm around his or her shoulders or provide other calming contact (Fig. 2). Data from several thousand postconflict observations in chimpanzees indicate that consolation reduces the recipient's arousal and follows the same sex difference as reported for sympathetic concern in children, with female apes providing comfort more often than males (*13*). Bonobos express the same tendency sociosexually by means

Living Links, Yerkes National Primate Research Center; and Department of Psychology, Emory University, Atlanta, GA 30322, USA. E-mail: dewaal@emory.edu

of genital contacts. A comparison of chimpanzee and bonobo brains supports the general view of bonobos as more empathic (*6*). This species has more gray matter in brain regions involved in the perception of distress, including the right dorsal amygdala and right anterior insula, and a better developed circuitry for inhibiting aggression (*14*).

Human empathy has been described as taking the perspective of another or imagining oneself in another's position. Psychologists commonly apply this cognitively demanding explanation of empathy even if the immediacy of the response hints at simpler processes. If we see a child fall and scrape its knee, we flinch, and exclaim "ouch!" as if what happened to the child happened at the same instant to ourselves. Another way of looking at empathy is as a multilayered phenomenon that starts with automatic state-matching based on motor mimicry and shared neural representations (*15*). We should not be surprised, therefore, by unconscious empathy, such as when human study participants mimic observed facial expressions and report corresponding emotions even though the expressions were presented too briefly for conscious perception (*16*). In this view, which may map onto nested neural processing (*17*), cognitive perspective-taking is a secondary development built around more elementary mechanisms, such as state-matching and emotional contagion (Fig. 3).

The evolution of empathy is thought to go back to mammalian maternal care. Whether a mouse or an elephant, a mother needs to be exquisitely in tune with indications of hunger, danger, or discomfort in her young. Sensitivity to emotional signals confers clear adaptive value. This hypothetical origin of empathy would explain the observed sex differences as well as the stimulating effect of oxytocin (*18*). That empathy is rooted in bodily connections between individuals is reflected in pain contagion in mice (*19*) and yawn contagion in apes and humans (*20*). Mirror neurons are often mentioned in this context, even though their precise role remains a point of speculation. The fact that these neurons were discovered not in humans, but in monkeys, supports the idea of evolutionary continuity. Social animals need to coordinate travel, communicate about danger, and assist group mates in need. Bodily synchronization and sensitivity to the emotional states of others ranges from rapid spreading of alarm through an entire group to a mother ape returning to a whimpering youngster to help it from one tree to the next by draping her body between the two. The first is a reflex-like transmission of fear, whereas the mother ape is more discriminating because she needs to assess the reason for her offspring's distress in order to ameliorate its situation.

The idea that empathy translates into altruism and helping is widely assumed for humans and has also been proposed for other mammals (*21*). Reports of spontaneous assistance among primates are abundant and are also available for elephants and cetaceans. For example, a female chimpanzee may react to the screams of her closest associate by defending her against an aggressive male, thus taking great risk on her behalf. Such coalitions are among the most systematically studied forms of cooperation in primatology (*22*). One advantage of an empathy-based explanation is its ability to explain "unrepaid" altruism, such as that shown toward nonreciprocating nonkin. This type of behavior is well illustrated by the adoption of orphans by wild male chimpanzees, who may devote years of costly care to unrelated juveniles (*23*). Although empathy, such as between a mother and offspring or between cooperation partners, is likely to be adaptive, not each and every application of this capacity needs to be for it to retain overall adaptive value.

Increasingly, the importance of mammalian prosocial tendencies is backed by experiments that range from demonstrating that rats give priority to the liberation of a trapped companion over eating chocolate (*24*) to those showing that apes are prepared to assist others even in the absence of incentives, go out of their way to give others access to food, or choose shared benefits over selfish ones (*25, 26*). Studies have also demonstrated more complex expressions of empathy among apes, such as "targeted helping" (assistance based on an appreciation of the other's specific needs), both in their spontaneous behavior (*21*) and during controlled experiments (*27*). This increased knowledge suggests that nearly the full spectrum of empathy-based altruism may be represented among nonhuman primates, including the cognitive perspective-taking that marks human altruism.

Empathy automatically produces a stake in another's welfare; that is, the behavior comes with

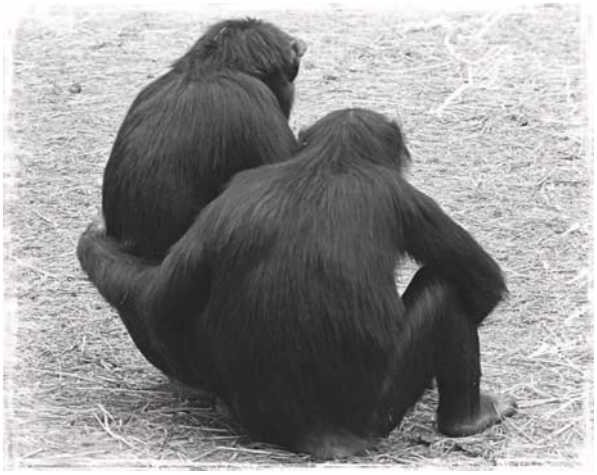


Fig. 2. Contact comfort is critically important in the lives of apes, such as here between two chimpanzees watching a disturbance in their group (photograph by Frans de Waal). Apes go out of their way to console distressed parties, showing the same sex difference in this tendency as humans.

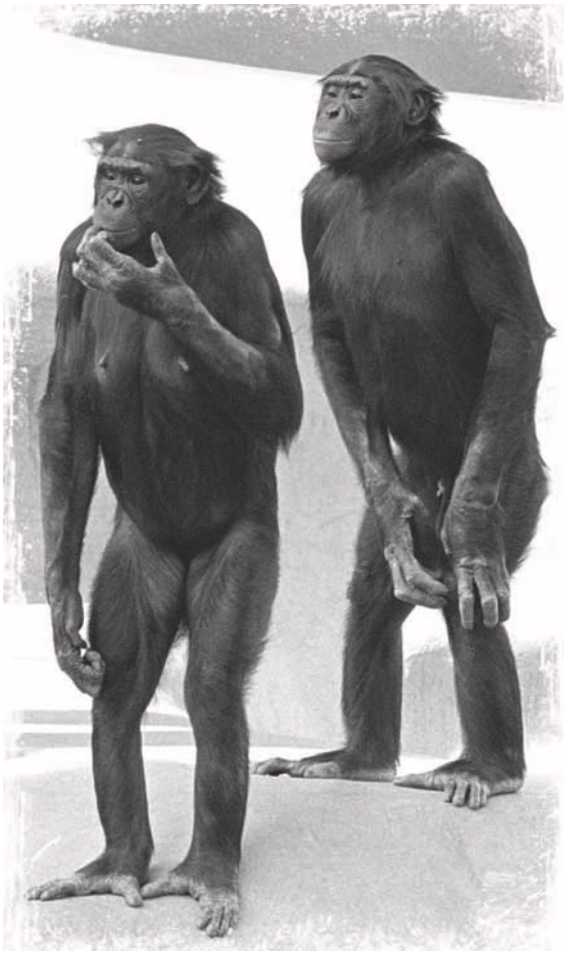


Fig. 1. Two bipedal bonobos, an adult female (left) and adolescent male (right), show the species' relatively long legs, which makes it anatomically more similar to early hominins than to the chimpanzee, with its longer arms and shorter legs (photograph by Frans de Waal). *Ardipithecus ramidus* may be the closest comparison to the bonobo in terms of its overall body proportions, grasping feet, and reduced canine teeth. *Ardipithecus* is thought to have been relatively peaceful, and bonobos are likewise marked by high sensitivity to others and low levels of violence.

an intrinsic reward, known as the "warm glow" effect. Humans report feeling good when they do good and show activation of reward-related brain areas (*28*). It will be important to determine whether the same self-reward system extends to other primates. We do know from studies on rodents, apes, and humans that empathy is biased toward the ingroup. For example, while watching the yawns of videotaped conspecifics, chimpanzees frequently joined the yawns of their own group members but not those of unfamiliar individuals (*23*). This ingroup bias makes sense from an evolutionary perspective, because it is with the members of one's own group that apes cooperate. At the same time, however, it poses a profound challenge for the modern human world, which seeks to integrate a multitude of groups, ethnicities, and nations. The flip side of the ingroup bias in empathy is lack of empathy for the outgroup, as is typical of xenophobia.

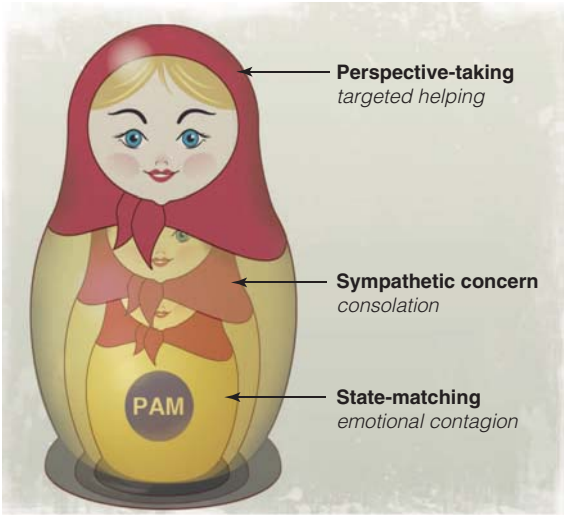


Fig. 3. The Russian doll model of multilayered empathy. The doll's inner core consists of the perception-action mechanism (PAM) that underlies state-matching and emotional contagion (*15*). Built around this hard-wired socioaffective basis, the doll's outer layers include sympathetic concern and targeted helping. The complexity of empathy grows with increasing perspective-taking capacities, which depend on prefrontal neural functioning, yet remain fundamentally connected to the PAM. A few large-brained species show all of the doll's layers, but most show only the inner ones.

Nevertheless, empathy may be our only hope to deal with these issues. We know that it can be activated by outsiders, even by members of a different species, such as when we empathize with a stranded whale and move it back into the ocean. This is not an outcome for which empathy evolved, yet once in existence, capacities are often emancipated from their evolutionary origin. If it weren't for empathy with all life forms, including enemy lives, soldiers would have no reluctance to kill nor would they return from the battlefield with PTSD. Although it is true that empathy has trouble reaching beyond the ingroup, it is an automated response that does not allow itself to be fully suppressed by rationalizations and political indoctrination. This is another lesson from World War II, with examples such as Oskar Schindler and the guardians of Anne Frank. To better understand the power of empathy requires investigation of its neurological basis as well as its evolutionary antiquity.

References

1. D. Grossman, *On Killing: The Psychological Cost of Learning to Kill in War and Society* (Back Bay Books, New York, 1995).
2. I. J. N. Thorpe, *World Archaeol.* **35**, 145 (2003).
3. D. P. Fry, *The Human Potential for Peace* (Oxford Univ. Press, New York, 2006).
4. R. W. Wrangham, L. Glowacki, *Hum. Nat.* **23**, 5 (2012).
5. C. O. Lovejoy, *Science* **326**, 74 (2009).
6. F. B. M. de Waal, *Bonobo: The Forgotten Ape* (Univ. of California Press, Berkeley, CA, 1997).
7. H. J. Coolidge, *Am. J. Phys. Anthropol.* **18**, 1 (1933).
8. B. B. Smuts, *Sex and Friendship in Baboons* (Harvard Univ. Press, Cambridge, MA, 1999) (originally published in 1985).
9. F. B. M. de Waal, A. van Roosmalen, *Behav. Ecol. Sociobiol.* **5**, 55 (1979).
10. F. B. M. de Waal, *Science* **289**, 586 (2000).
11. R. Boyd, *Science* **314**, 1555 (2006).
12. K. E. Langergraber, J. C. Mitani, L. Vigilant, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7786 (2007).
13. T. Romero, M. A. Castellanos, F. B. M. de Waal, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 12110 (2010).
14. J. K. Rilling *et al.*, *Soc. Cogn. Affect. Neurosci.* **7**, 369 (2012).
15. S. D. Preston, F. B. M. de Waal, *Behav. Brain Sci.* **25**, 1, discussion 20 (2002).
16. U. Dimberg, M. Thunberg, K. Elmehed, *Psychol. Sci.* **11**, 86 (2000).
17. J. Panksepp, *Science* **334**, 1358 (2011).
18. P. J. Zak, A. A. Stanton, S. Ahmadi, *PLoS ONE* **2**, e1128 (2007).
19. D. J. Langford *et al.*, *Science* **312**, 1967 (2006).
20. M. W. Campbell, F. B. M. de Waal, *PLoS ONE* **6**, e18283 (2011).
21. F. B. M. de Waal, *Annu. Rev. Psychol.* **59**, 279 (2008).
22. A. H. Harcourt, F. B. M. de Waal, *Coalitions and Alliances in Humans and Other Animals* (Oxford Univ. Press, Oxford, 1992).
23. C. Boesch, C. Bolé, N. Eckhardt, H. Boesch, *PLoS ONE* **5**, e8901 (2010).
24. I. Ben-Ami Bartal, J. Decety, P. Mason, *Science* **334**, 1427 (2011).
25. F. Warneken, B. Hare, A. P. Melis, D. Hanus, M. Tomasello, *PLoS Biol.* **5**, e184 (2007).
26. V. Horner, J. D. Carter, M. Suchak, F. B. M. de Waal, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 13847 (2011).
27. S. Yamamoto, T. Humle, M. Tanaka, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 3588 (2012).
28. W. T. Harbaugh, U. Mayr, D. R. Burghart, *Science* **316**, 1622 (2007).

10.1126/science.1220999

PERSPECTIVE

WARRIORS, LEVELERS, AND THE ROLE OF CONFLICT IN HUMAN SOCIAL EVOLUTION

Samuel Bowles

The origins of such varied features of contemporary life as the national state and the desire to uphold generous and civic social norms are to be found in a combination of conflict between groups and attenuation of both inequalities and conflicts within groups. In contrast to the adoption of a better tool or a more productive crop, which can be adopted by a single individual, a new institution works only if most people adopt it. This explains why collective action against those benefitting from the status quo at the expense of others, as well as conflict between groups governed by different norms and institutions, figures so prominently in our capacity to adapt to changing circumstances and to harness new knowledge for human benefit.

Conflict has a bad name, one that it richly deserves for the suffering, tragedy, and waste of human and material resources that it brings about. But conflict—both violent and civil, both within and between societies—has also been a midwife for humanity's most cherished values and institutions: among them democracy, the rule of law, and a propensity to help others and to abhor injustice.

I will make the case that it was warfare that culled Europe's once-motley collection of governments to produce the modern national state, which, as a result of subsequent conflicts within nations, would become liberal and eventually democratic. This occurred because, not content to free ride on the sacrifices of others, people were willing to take mortal risks in pursuit of democratic and liberal values. And this, if I am right, is itself a result of millennia of conflict between groups of ancestral humans where, Charles Darwin wrote, the groups with large numbers of "courageous, sympathetic and faithful members, who were always ready to...aid and defend each

Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA; and University of Siena, Siena, Italy. E-mail: samuel.bowles@gmail.com

other...would spread and be victorious over other tribes" (*1*).

Conflict and the Liberal Democratic State

Seven centuries ago, in what is now Italy, there were more than 200 distinct independent governing entities. Europe was governed by about 500 sovereign bodies: "empires, city states, federations of cities, networks of landlords, religious orders, leagues of pirates, warrior bands" (*2*). By World War I, fewer than 30 remained. A single political form had survived: the national state, a centralized bureaucratic structure maintaining order over a defined territory, with the capacity to mobilize substantial resources by taxation and borrowing and to deploy permanent armed forces.

What explains the competitive success of this novel form of rule? The simple answer is that national states won wars. An equally dramatic conflict-driven culling process took place in China between the fifth and third centuries BCE (*3*) and may also account for the first emergence of states not only in China but also in Mesopotamia, Mesoamerica, Peru, Egypt, and the Indus Valley (*4*). In Europe, success in warfare required mobilizing a willing or, at worst, compliant population. A system of taxation and military recruitment, coupled with the capacity to borrow large sums, made the difference, allowing rulers of national states to make war without resort to the unpopular ad hoc requisitioning of food, weapons, manpower, and animals (*2*).

All of this required a flourishing economy, the availability of credit, tax compliance, and the willingness to serve rulers in war. These, in turn, were fostered by the diffusion of civic norms—voluntary tax compliance, willingness to risk danger in war for a ruler or nation, and respect for property rights—which, although costly to the individual, were essential to a nation's success in war.

In part as a result of its success in Europe, replicas of the national state were exported, often at gun point, but also by emulation on the part of those seeking to preserve their own autonomy. The European model of government—often in highly authoritarian form, as in the colonies—flourished throughout the world, extinguishing competing forms of organization. With the national liberation wars and independence movements of the 19th and 20th centuries, together with subsequent social movements for expanded suffrage and civil rights, many of these states, too, would become democratic.

Some kinds of progress avoid the tragedies of war and civil strife: A more efficient energy source or an advance in personal hygiene comes along, and those who adopt it profit as a result. But the main dynamic of social norms and institutions has a different logic. A novel system of property rights, governance, or marriage, or a new medium of exchange or of communication, only works when it is widely adopted. These systems are termed conventions. Switching from one to the other is known as a coordination problem and, as the term suggests, this occurs through collective, not individual, action when the number of people rejecting the status quo is sufficient to tip the population to an alternative convention. A new convention is not something that you can opt out of, and it is often the powerful and wealthy in the status quo convention who will be the losers in the new. So it is no surprise that shifting from one to another generates conflict, whether violent or civil. This is why strikes, demonstrations, and wars provide so many of the punctuation marks of history (along with new technologies).

The eventual democratization of the national state exemplifies just this process. American high school students are taught that their democratic constitution was the gift of the landed and commercial elites of the 13 former colonies. James Madison and the other authors of *The Federalist Papers,* the story goes, convinced the haves that the have nots would never be able to unite sufficiently to redistribute wealth. The elites could safely take a chance on democracy. But that is just one of America's national myths. The United States would wait more than a century and a half to meet the elementary standard of democratic rule by extending suffrage to virtually all adults (with the Voting Rights Act of 1965), a process propelled by the victories of abolitionists, slaves and their descendants, workers, and women demanding the vote (Fig. 1, top).

Elsewhere, conflict played an even more critical role in the advance of democracy (*5*). With the exception of New Zealand, universal suffrage was not won anywhere until the 20th century, and elites rarely conceded it without a fight (*6*). Representative institutions with limited voting rights came first in Europe and its global offshoots, often as a result of the defeat of a landed elite, as in France. This was followed, in most cases much later, by the equally contentious extension of the vote. World War I sent millions of disenfranchised soldiers to their graves; in its course and immediate aftermath, nine European nations extended the vote to all males, most granting the vote to women at the same time (Fig. 1, bottom).

Democracy has belatedly come to El Salvador, South Africa, and many of the former Communist Party–ruled nations, but only because peasants, workers, and other ordinary citizens were willing to risk jail and much worse (*7*–*9*). A similar process may now be under way, if haltingly, in the Arab world. Conflict and elites' attempts to forestall conflict were no less essential to the eventual adoption of policies to ensure the modicum of equality of opportunity and social insurance that most citizens of liberal democracies now take for granted.

Cooperation and Conflict in Prehistory

All of this required collective action on the part of those excluded from the political process who, like the youth of Tahrir Square, were willing to sacrifice for others beyond the immediate family. The fact that humans, uniquely among animals, are like this is itself arguably the result of millennia of conflict between groups of ancestral humans, as Darwin wrote. Some have doubted Darwin's account because for most animals, gene flow (due to migration) would minimize genetic differences between groups, and hence nullify the genetic effects of group competition (*10*). But recent evidence suggests that this may not be the case for a number of species (*11*–*13*), including recent human foragers, whose population structures may resemble those of our Late Pleistocene ancestors (*14*). Others may have doubted Darwin's conflict-based account because they believe warfare to be a postagricultural revolution corruption of our naturally peaceful disposition. But hunter-gatherer burials with smashed skulls, broken or missing forearms (taken as trophies), and stone points embedded in bones tell a different story, as does ethnographic evidence that warfare was a leading cause of death among some recent hunters and gatherers (*14*).

I have shown (*15*, *16*) that we can plausibly infer from these data that the degree of mortal conflict and extent of genetic differences among ancestral forager groups were jointly sufficient to have allowed the evolution of a genetically transmitted predisposition to contribute to common projects (including defense and predation vis a vis other groups), even when one's individual fitness would have been enhanced by free riding on those who would "aid and defend each other." Whatever the balance of cultural and genetic factors in the evolution of human cooperativeness, between-group conflict almost certainly played a pivotal role. Related empirical and theoretical results are consistent with this view (*17*–*18*).

Warriors and Levelers

This explanation of a warlike provenance of human altruism comes with an interesting twist: In addition to making war, our hunter-gatherer ancestors almost certainly built institutions to share food and information, to make decisions by consensus, and to gang up on would-be dominants or free riders who would monopolize reproductive and material resources or exploit the cooperation of others (*19*–*21*). These practices, called reproductive leveling, reduced within-group differences in material wealth and reproductive success, resulting in a less tilted evolutionary playing field and thereby giving the altruistically inclined a better chance of survival. The result, paralleling natural selection and affecting its course, was a prehistoric culling of institutions not unlike that which produced the national state in early modern Europe. In simulations of this gene-culture coevolutionary process in prehistoric populations, when these leveling processes are prevented from evolving, natural selection produces a self-interested species (except under empirically implausible

Fig. 1. (**Top**) Battle of Antietam, 17 September 1862, during the U.S. Civil War that ended slavery. (**Bottom**) British Chartists demonstrating at Kennington Common on 10 April 1848. Enactment of their demand for universal male suffrage would not come about until 70 years later as World War I drew to a close.

parameter values) just as the group selection skeptics predicted (*22*).

The great shake-out of European protostates favored a similar kind of leveling. In many of the states that survived the winnowing process, the rule of law limited the predations on the weak by the strong. As celebrated in a House of Commons speech in 1763 attributed to William Pitt the Elder, "The poorest man in his cottage may bid defiance to all the forces of the Crown. It may be frail—its roof may shake—the wind may enter—the rain may enter—but the king of England cannot enter—all his force dares not cross the threshold of the ruined tenement" (*23*). Eventually, democratic rule and the organization of trade unions allowed the have nots, like their distant forager ancestors, occasionally to curb those claiming too large a share of the economic pie. Thus, not unlike the Late Pleistocene, the institutional selection that produced the modern liberal state was driven by processes attenuating inequality and conflict within nations, working in combination with conflict between nations. The modern nationalistic welfare state (e.g., France) is a result of this evolutionary process.

The biologists John Maynard Smith and Eörs Szathmary proposed that the major transitions in biology—for example, the emergence of multicellular organisms—occurred when competition within entities was suppressed (*24*). There may be a similar process at work among our forager ancestors and in early modern Europe: Success in competition between entities is more likely when competition and conflict within entities is moderated (*15*, *19*, *25*, *26*). In a paper on slime mold, Frank writes (*27*): "...competition among lower-level units is suppressed in the formation of higher-level evolutionary units, ...mutual policing and enforcement of reproductive fairness are also required for the evolution of increasing social complexity."

Unlike multicellular organisms, forager bands and nations cannot ensure their competitive effectiveness the way slime molds do it: by suppressing the autonomy of the lower-level entities making them up. Instead, for human groups, prevailing in intergroup contests requires cooperation among individuals. If the above account is correct, this is fostered both by an altruistic predisposition among group members and by reproductive leveling, the rule of law, democracy, and other practices that limit the extent to which leaders take what is considered to be unfair advantage over others.

However, there is nothing intrinsic to warfare that guarantees similar outcomes. For all their cooperativeness and reproductive leveling, forager bands were no match for the hierarchical agrarian and later industrial states that decimated them. In the two cases here—forager bands and national states—a combination of within-group cooperation and leveling appears to have contributed to success in between-group conflict, and hence was favored. But the logic of competition among differing institutions and social norms does not guarantee such benign results anymore than natural selection maximizes the fitness of a species or competition among profit-seeking firms results in an efficient allocation of economic resources.

Thus, war is hardly sufficient for the evolution of altruism or leveling. Nor is war necessary. People routinely act generously in workaday situations. Natural disasters often bring out the best in us and inspire heroic sacrifices on behalf of others. During the Late Pleistocene, groups of cooperating foragers would have been more likely to survive not only challenges by other groups but also the extraordinary climatic shocks of that period. Could not human cooperativeness and leveling have thus evolved in the absence of between-group conflict? It could have, but, in light of the evidence, I doubt that it did.

Legacy and Destiny

It seems likely, too, that conflict will remain important for human progress. But does this require the violence, suffering, bigotry, and waste characteristic of the conflicts of the past along with the cultural inheritance of this dismal trajectory, an unpleasant nexus of predispositions that Choi and I call "parochial altruism," marked by gen-

erosity toward those we call "us" and hostility and intolerance toward "them" (*16*)?

I do not think so: Our legacy need not be our fate. We could not have become what Gintis and I call a cooperative species (*28*) were we not, *par excellence*, a cultural animal. Among the lessons of our past are not only the grisly truths on which I have dwelled but also the fact that our us's and them's are not primordial. On world historic time scales, we make and unmake these pronouns of exclusion at lightning speed. For ancestral humans, making peace was no less essential than surviving wars [as Boehm points out in his contribution to this issue (*29*)].

The unsung virtue of European and many other forms of nationalism is that it obliterated the hundreds of petty us's and them's that once divided valley from valley, dialect from dialect, and even neighborhood from neighborhood (*30*, *31*). The tricolor, the stars and stripes, and the other national banners have not, of course, put an end to intolerance and bigotry within nations. But the willingness of voters to elect members of groups whom they recently despised, exploited, fought, or enslaved, and to pay taxes to extend economic opportunity and a modicum of security to once-excluded peoples is testimony to the fragility of the parochial aspects of altruism.

Nationalism helped convince once-warring peoples—Protestant and Catholic, Florentine and Roman—to bury the hatchet, if not their differences. Paradoxically, globalism may carry a similar process across national boundaries. The parochial face of nationalism itself may be softened by the globalization of interpersonal contact and concern, now facilitated by the shrinking of space. And if, as seems likely, democracy should continue to spread, relations among nations may come to reflect what political scientists call the democratic peace (*32*) and follow the less bellicose avenues of economic and cultural competition and emulation.

References and Notes

1. C. Darwin, *The Descent of Man* (D. Appleton and Company, New York, 1873).
2. C. Tilly, *Coercion, Capital, and European states, AD 990-1990* (B. Blackwell, Cambridge, MA, 1990).
3. V. Hui, *War and State Formation in Ancient China and Early Modern Europe* (Cambridge Univ. Press, Cambridge, 2005).
4. C. S. Spencer, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 7119 (2010).
5. B. J. Moore, *Social Origins of Dictatorship and Democracy; Lord and Peasant in the Making of the Modern World* (Beacon Press, Boston, 1966).
6. G. Therborn, *New Left Rev.* **103**, 3 (1977).
7. E. J. Wood, *Forging Democracy From Below: Insurgent Transitions in South Africa and El Salvador* (Cambridge Univ. Press, Cambridge, 2000).
8. M. Beissinger, *Nationalist Mobilization and the Collapse of the Soviet State* (Cambridge Univ. Press, Cambridge, 2002).
9. R. Collier, *Paths Toward Democracy: The Working Class and Elites in Western Europe and South America* (Cambridge Univ. Press, Cambridge, 1999).
10. G. C. Williams, *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought* (Princeton Univ. Press, Princeton, NJ, 1966).
11. D. S. Wilson, E. O. Wilson, *Q. Rev. Biol.* **82**, 327 (2007).
12. E. O. Wilson, B. Hölldobler, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 13367 (2005).
13. T. L. Goldberg, R. W. Wrangham, *Anim. Behav.* **54**, 559 (1997).
14. S. Bowles, *Science* **324**, 1293 (2009).
15. S. Bowles, *Science* **314**, 1569 (2006).
16. J.-K. Choi, S. Bowles, *Science* **318**, 636 (2007).
17. L. Lehmann, M. W. Feldman, *Proc. Biol. Sci.* **275**, 2877 (2008).
18. H. Bernhard, U. Fischbacher, E. Fehr, *Nature* **442**, 912 (2006).
19. C. Boehm, *Moral Origins: Social Selection and the Evolution of Virtue, Altruism, and Shame* (Basic Books, New York, 2012).
20. P. Wiessner, in *Food and the Status Quest: An Interdisciplinary Perspective* P. Wiessner, W. Schiefenhovel, Eds. (Berghahn, Oxford, 1982).
21. P. Wiessner, *Hum. Nat.* **16**, 115 (2005).
22. S. Bowles, J. K. Choi, A. Hopfensitz, *J. Theor. Biol.* **223**, 135 (2003).
23. J. Bartlett, *Familiar Quotations* (Little Brown, Boston, ed. 16, 1992).
24. J. M. Smith, E. Szathmary, *The Major Transitions in Evolution* (Oxford Univ. Press, Oxford, 1995).
25. I. Eibl-Eibesfeldt, *J. Comparative Ethology* **60**, 177 (1982).
26. C. Boehm, *J. Soc. Biol. Struct.* **5**, 413 (1982).
27. S. A. Frank, *Nature* **377**, 520 (1995).
28. S. Bowles, H. Gintis, *A Cooperative Species: Human Reciprocity and its Evolution* (Princeton Univ. Press, Princeton, 2011).
29. C. Boehm, *Science* **336**, 844 (2012).
30. E. Gellner, *Nations and Nationalism: New Perspectives on the Past* (Cornell Univ. Press, Ithaca, 1983).
31. E. Weber, *Peasants into Frenchmen: The Modernization of Rural France, 1870-1914* (Stanford Univ. Press, Stanford, 1976).
32. A. Dafoe, *Am. J. Pol. Sci.* **55**, 247 (2011).

Ackowledgments: The author declares no competing interests and would like to thank the Behavioral Sciences Program at the Santa Fe Institute for funding; R. Allen, C. Boehm, R. Boyd, J. Cohen, S. Naidu, P. Richerson, J. Roemer, J. Sabloff, N. Sambanis, and E. Wood for valuable comments; and his teacher, the late B. Moore Jr., who sparked his interest in the topic.

10.1126/science.1217336

REVIEW

LIFE WITHOUT WAR

Douglas P. Fry[1,2]

An emerging evolutionary perspective suggests that nature and human nature are less "red in tooth and claw" than generally acknowledged by a competition-based view of the biological world. War is not always present in human societies. Peace systems, defined as groups of neighboring societies that do not make war on each other, exist on different continents. A comparison of three peace systems—the Upper Xingu River basin tribes of Brazil, the Iroquois Confederacy of upper New York State, and the European Union—highlight six features hypothesized to be important in the creation and maintenance of intersocietal peace: (i) an overarching social identity, (ii) interconnections among subgroups, (iii) interdependence, (iv) nonwarring values, (v) symbolism and ceremonies that reinforce peace, and (vi) superordinate institutions for conflict management. The existence of peace systems demonstrates that it is possible to create social systems free of war.

War—a group activity involving lethal aggression between communities—and other forms of violent conflict occur all too regularly in the 21st century and contribute substantially to human suffering. At the same time, most daily human behavior, within and across societies, is nonviolent. Conflict—defined generally as perceived divergence of interests—occurs regularly within and between societies and can be handled in many ways, only a few of which involve any physical violence (*1*, *2*). With variation from one culture to the next, disputants, for example, may seek the help of an impartial mediator to resolve their disagreements, appear in court, negotiate the payment of compensation, or practice avoidance.

A New Perspective

A dominant evolutionary perspective, as captured in Tennyson's famous phrase "nature, red in tooth and claw," has proposed that competition, often in the form of violence, is the evolutionary norm (*3*–*7*). It appears, however, that this perspective may be shifting toward a new understanding that, although not totally dismissive of self-interested competition and conflict, nonetheless draws on recent advances in evolutionary theory (*3*–*5*) and a substantial body of human and nonhuman animal data (*7*, *8*) to show that cooperation, sharing, helping, and reconciliation also have a solid evolutionary basis (*3*–*11*).

Traditionally, warfare has been seen as ancient (*12*–*14*), but this view is also being recon-

[1]Peace, Mediation and Conflict Research, Department of Social Sciences, Åbo Akademi University in Vasa, Post Office Box 311, FIN-65100, Vasa, Finland. [2]Bureau of Applied Research in Anthropology, School of Anthropology, Post Office Box 210030, Tucson, AZ 85721–0030, USA. E-mail: dfry@abo.fi

sidered (*2*, *15*). Chimpanzee intergroup killings have been used to make inferences about past hominid behavior (*13*), but ancestral hominid conflict behavior may have been more bonobo-like—that is, nonlethal—than chimp-like (*3*), as perhaps reflected in the small, nonprojecting canines and minimal sexual dimorphism of 4.4-million-old *Ardipithecus ramidus* (*16*). Likewise, humans and their predecessors have long been characterized as hunters, but recently a reconsideration of several types of evidence suggests that they might more accurately have been considered prey (*10*).

Similarly, a much-discussed suggestion based on Yanomamö data, that warriors have higher reproductive success than nonwarriors (*17*), has been reevaluated theoretically, mathematically, and empirically with the conclusion that the opposite may actually be the case (*2*, *15*, *18*, *19*). First, in terms of theory, computer simulations indicate that unrestrained, escalated forms of aggression do not fare as well as strategies of limited agonism (*20*, *21*). Second, a mathematical reanalysis of the data on the Yanomamö men in the original study revealed that those reported to have killed averaged more offspring in part because they were more than 10 years older than the men who had never killed (*2*). Third, attempted replications using data on the Cheyenne and the Waorani failed to support the original findings; to the contrary, fitness was found to be negatively correlated with killing (*2*, *18*).

The evidence for a new orientation that gives cooperation and peacefulness a seat at the evolutionary table comes from a variety of sources. From ethology, studies document cooperation, empathy, and conflict resolution in various animal species (*3*, *7*–*9*). Mammalian patterns of intraspecific agonism correspond with game theoretic simulations (*20*), namely, escalated fighting among rival conspecifics is very rare compared with the widespread use of noncontact displays and ritualized contests (wherein serious injuries or death are unlikely), and similar patterns of restraint characterize much human conflict as well (*2*, *21*, *22*). Primates, including humans, readily cooperate and reconcile after aggression, especially within social groups when partners are mutually dependent on each other (e.g., as allies in dominance struggles) (*3*, *7*, *9*, *23*). A reevaluation of nomadic forager data—such as on the Yahgan of South America, the Saulteaux and Paiute of North America, the Semang and Vedda of Asia, and the Ju/'hoansi and Mbuti of Africa—shows that warfare is most often absent at this ancestral level of social organization (*2*), whereas cooperating and helping (e.g., the sharing of meat) occur without exception in nomadic band societies (*2*, *24*, *25*). Furthermore, neurobiological research shows that humans receive an immediate biochemical reward in oxytocin for cooperating (*26*), and this may be part of an evolved brain-reward system in humans related to cooperation, trust, and altruism (*26*). Finally, research from military science suggests that it is more difficult to get soldiers to kill in combat than commonly assumed. The resistance by soldiers to killing the enemy has been documented across various battlefield settings, for instance, among French officers in the 1860s, the battle of Gettysburg in the U.S. Civil War, soldiers from Argentina in the Falklands War, and U.S. troops during World War II (where it has been estimated that at most 25% of the combat soldiers shot at the enemy) (*19*, *27*). The relevant point, as General S.L.A. Marshall, who studied firing rates during World War II, concluded, is that "the average and normally healthy individual…[has] an inner and usually unrealized resistance toward killing" (*28*). Subsequently, this reluctance to kill has been largely overcome by the U.S. military through intense training in reflexive firing (*27*).

In sum, the traditional focus in the evolutionary sciences has been on competition and violent conflict (*5*, *7*); however, this perspective is shifting toward one that is more appreciative of cooperation, peacemaking, empathy, and sharing. Obviously, humans have the capacity to engage in war, but a growing body of studies on animals and humans suggests that nature is less violent than commonly has been assumed (*2*–*5*, *7*–*9*).

Nonwarring Societies

Ethnographically, most societies engage in warfare, but there are some that do not (*2*, *29*). Considering the existence of nonwarring societies is important because it demonstrates that humans are capable of living without war. Nonwarring societies can be found in various locations around the globe, for example, the Machiguenga swidden farmers of Peru (*30*) and the Batek of Malaysia (*31*). Among the Batek, core values, as evident in everyday social interaction, include helping anyone in need, respecting others, being noncompetitive and nonviolent, and sharing food (*31*). The Mardu Aborigines of Australia's Great Western Desert, who were studied by Tonkinson beginning in the 1960s, had remained up until that time relatively isolated and unaffected by outside influences. The Mardu and neighboring Aborigines do not practice warfare; even the concept is alien to the Mardu because their language lacks words for feud and war (*32*). The types of conflicts that arise, for instance elopements or sorcery accusations, tend to be interpersonal, and the Mardu routinely resolve such grievances when several bands gather together (*32*). Many additional cases of nonwarring societies have been described from the Pemon and Piaroa of South America, the Kawaiisu and Karok of North America, the Central and Copper Inuit of the Arctic, and the Ladakhi and Lepcha of Asia, to nearly all Australian Aborigine societies (*2*). It is also worth noting that some nations (for example, Iceland, Switzerland, and Sweden) have avoided wars for generations. Thus, war is not always and everywhere present.

One robust anthropological finding is that complexity of social organization correlates positively with warfare. Hierarchical societies such as chiefdoms, kingdoms, states, and empires are more likely to engage in war and practice more severe forms of warfare than are comparatively egalitarian tribes or highly egalitarian nomadic forager bands (*2*). In bands, especially, numerous factors protect against war. For example, individuals have close relatives in neighboring bands, lethal disputes generally have very personal, not political, causes (e.g., due to sexual jealousy, an insult, or revenge-seeking by a homicide victim's family), there are no caches of stored food or other goods to plunder, no one possesses the authority to command other band members to fight, and population densities tend to be very low with adequate resources spread over wide areas (*2*, *25*).

Table 1. Examples of peace systems. Peace systems are groups of neighboring societies that do not make war on each other (and sometimes not with outsiders either). Peace systems can be found in various parts of the globe.

Geographical location	Peace system
Australia	Peoples of the Great Western Desert (e.g., the Mardudjara/Mardu, Gugadji, Walmadjeri, and Pintupi, among numerous others)
Canada	Montagnais-Naskapi and East Main Cree of the Labrador Peninsula
India	Nilgiri Plateau/Hills societies (Toda, Kota, Badaga, and Kurumbas)
Malaysia	Central Peninsular Orang Asli societies (e.g., Batek, Jahai, Semai, Chewong, and Btsisi)
Greenland	Native Inuit populations
United States and Canada	Iroquois Great League of Peace (Cayuga, Mohawk, Oneida, Onondaga, Seneca, and Tuscarora)
Brazil	Ten tribes of the Upper Xingu River basin (Kuikuru, Kalapalo, Nafukuá, Matipú, Mehinaku, Wauja, Yawalapití/Yaulapití, Kamayura/Kamaiyura, Aultí, and Trumaí)
Europe	European Union (27 member countries and growing)

Peace Systems

As reflected in Table 1, peace systems—groups of neighboring societies that do not make war on each other and sometimes

not with outsiders either—have been documented on different continents (*33*–*35*). Some peace systems are strictly nonwarring, whereas other peace systems are only nonwarring within the system itself. For example, the Batek, Btsisi, Chewong, Jahai, Semai and other nearby Malaysian societies constitute a peace system comprised of cultures that totally shun war (*31*).

In addition to ethnographic cases, peace systems are reflected in the archaeological record. For example, very complete archaeological data exist for the prehistoric Anasazi of the southwestern United States, which clearly shows the time periods when war was either absent or present. War leaves tell-tale marks such as habitation sites protected by stockades, evidence of widespread fire or destruction, high percentages of violent death reflected in burials, and so forth (*2*). The archaeological record across the Anasazi cultural area from before 700 CE until almost 1200 CE shows no evidence of war (*36*). At the end of this period, as the climate changed to drought conditions, war appears. By the mid-13th century, in marked contrast to the preceding centuries, the evidence for war is unmistakable. Village destruction is evident, settlements have shifted to defensive locations, and lethal trauma is endemic in the skeletal population (*36*). After a successful 500-year run, the Anasazi peace system broke down, likely under the pressure of demographic and environmental stress.

An examination of existing peace systems can provide insights for creating peace in other settings (*34*). Preliminary comparisons suggest that, for some peaceful groups, nonwarring may be simply the behavioral default (passive systems), perhaps because war is traditionally unknown among the member societies, as seen in the Malaysian case mentioned previously, whereas other peace systems are more actively created and maintained (active systems).

A comparison of active peace systems suggests that common features that can be hypothesized to be important include (i) an overarching social identity, (ii) interconnections among subgroups, (iii) interdependence (ecological, economic, and/or defensive), (iv) nonwarring values, (v) symbolism and ceremonies that reinforce peace, and (vi) superordinate institutions and conflict management. Each of these features will be briefly considered using ethnographic examples from various cultures, but with a primary focus on three examples of active peace systems, the Upper Xingu River basin tribes in Brazil, the Iroquois Confederacy, and the European Union (EU) (Table 2).

Table 2. Features of five peace systems compared. Additional information on the three active peace systems is provided in the text. Information on the two passive peace systems is presented here for comparative purposes.

Feature	Active peace systems			Passive peace systems	
	Upper Xingu River basin	Iroquois Great League of Peace	European Union	Central Malaysia Orang Asli	Montagnais-Naskapi and East Main Cree
Overarching social identity	Yes	Yes	Growing	No	No
Interconnections among subgroups	Yes	Yes	Yes	?	?
Interdependence	Augmented	Augmented	Pivotal	No	No
Values for peace	Yes	Yes	Yes	Yes (nonviolence)	Not reported
Symbols, rituals, ceremonies for peace	Yes (chiefly ceremonies, peace myths)	Yes (Tree of Peace, rituals of condolence)	Yes (flag, anthem, Euro currency)	No	No
Intergroup conflict management	Weak: harangues, wrestling	Strong: Great Council of Chiefs, compensation	Intermediate: EU Court, Commission, Parliament	Weak: avoidance and toleration	Weak: avoidance, toleration, public opinion
Overarching governance	No	Yes (Council of Chiefs)	Yes (e.g., EU Parliament, Commission)	No	No

All three of these peace systems have eliminated warfare within the system, but not necessarily against outsiders. The 10 Upper Xingu River basin tribes protected themselves if attacked by aggressive outsiders (*37*). Additionally, and as in many societies, homicide occasionally occurred, but homicide is not war. Among the original Iroquoians—the Mohawk, Oneida, Onondaga, Cayuga, and Seneca nations—archaeology and ethnohistory clearly document chronic feuding, warring, and cannibalism before the creation of the peace system put an end to the carnage within the new confederacy (*38*, *39*). Since its beginning in the second half of the 15th century, the Iroquois Confederacy, also known as the League of Peace, "proved remarkably durable, maintaining the peace among the Iroquois for over three hundred years" (*39*). EU member nations have contributed troops to recent wars outside the EU borders, but the use of military force within the EU has not occurred for more than 60 years. In post–World War II Europe, as the horrific memories of widespread death, bombings, blackouts, food rationing, hunger, concentration camps, and mass graves were still fresh in the minds of the survivors, the motivation was very strong to devise a way to prevent future European wars. Thus, creating a sustainable peace was the primary driving force behind European integration. These three cases do not show a total abstention from war, but rather they illustrate how clusters of neighboring societies have successfully created peace systems among themselves.

Identity: Expanding the Us

The promotion of an "us-versus-them" mentality can facilitate intergroup hostility (*1*, *40*); however, at least some successful peace systems form a common identity that helps to promote peace. For example, among the 10 tribes of the Upper Xingu River basin, which represent four different language groups, identities go beyond individual tribal membership (*34*, *37*, *41*). The societies have "expanded the us" to encompass a common identity with the other tribes.

The Iroquois also expanded the us and metaphorically referred to their confederacy as a longhouse, symbolically denoting an extended family living together in peace (*42*). "They sought to expand their League of Peace, and to embrace ever more people as kinsmen who would share the peace, prosperity, and security of the Iroquois Longhouse" (*38*). An evolving pan-Iroquois identity is reflected in many ways. As the peace system developed, the previous practice of exacting blood revenge over a homicide was replaced by the payment of compensation; the former practice of cannibalism within the system became obsolete as outsiders became insiders; the distinct pottery styles of past eras became progressively uniform across the region, reflecting a common identity; intermarriage increased, being simultaneously a cause and a result of the expanding Iroquois social identification; ritualized adoptions turned nonkin into relatives, and importantly, the use of kinship imagery and terminology supported the unifying view of all Iroquoians as relatives (*38*, *39*).

Within the EU, a new overarching identity is emerging (*39*). This is evident in the issuance of EU passports, EU automobile license plates, the

Euro as a common currency (in most member countries), the free movement of EU citizens across borders, democratic elections for EU parliamentarians, and an EU flag and anthem (*34*). In short, the trend is toward a new pan-European identity that parallels how the Upper Xingu and the Iroquois peoples developed additional overarching social identities. Expanding the us is a powerful force in the service of peace (*33*, *39*).

Intergroup Ties

Intergroup bonds of friendship and kinship discourage violence (*1*, *33*). For example, as we have seen, the Mardu of Australia do not feud or war. The Mardu bands are interlinked with each other and also with bands from neighboring societies as part of a larger kinship system, and cross-cutting ties help to promote peaceful relations among the groups (*35*). The intergroup ties are diverse. This is due in part to the numerous interconnections (shared values, norms, religion, cosmology, friendship, kinship, and marriage alliances) that link all Mardu bands with each other. Such commonalities and linkages greatly facilitate the resolution of disputes (*32*). Similarly, among the nonwarring Ju/'hoansi of Africa, everybody has kin and trade partners in other groups and, as among the Mardu, intergroup connections discourage intergroup hostility (*2*). Intermarriage also served a similar peace-sustaining function for the Iroquois and Upper Xingu River peace systems (*33*, *37*, *38*, *41*). The emergence of a pan-Iroquois material culture reflects the progressive social integration across the confederacy (*38*). The general principle is that the existence of cross-cutting ties such as ceremonial unions, fictive and consanguineal kinship, economic partnerships, and friendships, decreases the chances that conflicts will result in war.

Interdependence as a Key Factor

Interestingly, data from disciplines as diverse as primatology, anthropology, social psychology, and political science converge on showing the importance of interdependence for promoting cooperation to achieve a superordinate goal (*2*, *32*, *40*, *43*), although economic interdependence may play a more important role later in the process (*39*). In turn, engaging in cooperation is beneficial for relationships and thus can contribute to peaceful intergroup relations (*2*, *34*, *40*, *43*). There are various types of interdependence, some imposed by circumstance and some purposefully created. The harshness of the physical environment can be a unifying force. The solution in Africa's Kalahari Desert or Australia's Great Western Desert is for local groups to reciprocally allow each other access to water and food resources (*2*, *33*). To let disputes harden into feuding or warfare under such conditions of ecological interdependence would be suicidal (*2*, *32*).

Fig. 1. The Tree of Peace. The insignia on the flag of the Oneida Nation, one of the original five member societies of the Iroquois Confederacy, depicts the Tree of Peace. The Cayuga, Mohawk, Oneida, Onondaga, and Seneca (later joined by the Tuscarora) brought an end to the chronic warfare among themselves, represented ritualistically as they interred their war hatchets and war clubs under the Tree of Peace. An underground river washed all their weapons away, making them irretrievable, thus reflecting symbolically the five nations' commitment to unity and peace. The eagle of vigilance is shown, as are the white roots for peace, the latter symbolizing the Iroquois desire to spread peace to all the peoples of the world. The legendary prophet Deganawidah proclaimed as the weapons of war were eliminated, "We have rid the earth of these things of an Evil Mind…. Thus…shall the Great Peace be established, and hostilities shall no longer be known between the Five Nations, but peace to the United People" (*42*).

The tribes of the Upper Xingu peace system enhance interdependence by specializing in the production of particular trade goods, such as pottery, hardwood bows, or salt, and this type of specialization creates multiple economic interdependencies among them (*37*). The founders of the EU augmented interdependence as part of a deliberate plan to create a new level of governance. An explicit impetus behind European integration was to eliminate the threat of war in the region (*34*, *44*). The creative approach was to build economic and political interdependence by incrementally integrating the national economies (*34*). The first step, taken in the 1950s, was placing coal and steel—critical resources in times of peace and war—under supranational control. Thus began an agenda of cooperation and unification (*44*). Current economic challenges within the EU, such as high unemployment and debt burden in some countries, in no way diminish the success of European integration and unification in creating a continent safe from war. In Europe today, less than 70 years since World War II ravaged the continent, war among EU member nations has become highly unlikely (*34*). Using interdependence to deliberately create a regional peace system is a remarkable achievement.

Values for Peace

Some value orientations are more conducive to peace than others (*35*). In the value system of the Upper Xingu tribes, the warrior role is shunned—peace is moral, but war is not (*37*, *41*). In that values become internalized within the minds of people and serve as guidelines for behavior, the promotion of antiwar values in society has a role to play in sustaining the peace (*2*, *35*). The Iroquois made their value of peace within their confederacy explicit: "Thus we bury all the weapons of war out of sight, and establish the 'Great Peace.' Hostilities shall not be seen nor heard of any more among you, but 'Peace' shall be preserved among the Confederated Nations" (*38*).

The EU was founded with the explicit goal of bringing peace and prosperity to Europe. Peace-related values such as democracy, social equality, human rights, and respect for the law serve as the EU's moral compass, as explicitly stated: "Promoting these values, as well as peace and the well-being of the Union's peoples are now the main objectives of the Union" (*45*). The actualization of these values in the EU is reflected in numerous ways, including health care provided for all people, inexpensive (sometimes free) university education, a high standard of living, effective public transportation, strong democratic institutions, sufficient retirement security, affordable child care, paid parental leave, and so on (*46*). Rates of violent crime are much lower in the EU nations than in the United States (*2*). In short, a strong argument can be made that Europeans have successfully devised social institutions that promote not only peace but also respect for the law, justice, democracy, equality, and human rights among EU nations.

Symbolism and Ceremonies for Peace

Symbols and ceremonies reinforce unity and commitment to peace. All the Upper Xingu tribes participate in ceremonies to mourn the deaths of deceased chiefs and to inaugurate new ones. Joint ceremonies help to unify the tribes and reinforce their expanded shared identity as members of the same broader peaceful society. One Xinguano expressed the intention in this way: "We don't make war; we have festivals for the chiefs to which all of the villages come. We sing, dance, trade and wrestle" (*41*). Like the longhouse that symbolically represented the Iroquois Confederacy as one family, the Tree of Peace was a powerful symbol for peace and unity (Fig. 1). According to an often-recounted legend, at the formation of the League of Peace, the weapons of war were buried beneath the tree and then washed away into a subterranean cavern. Symbolically, the white roots of the Tree of Peace represent the desire for peace to spread beyond the confederacy to embrace neighboring societies in all di-

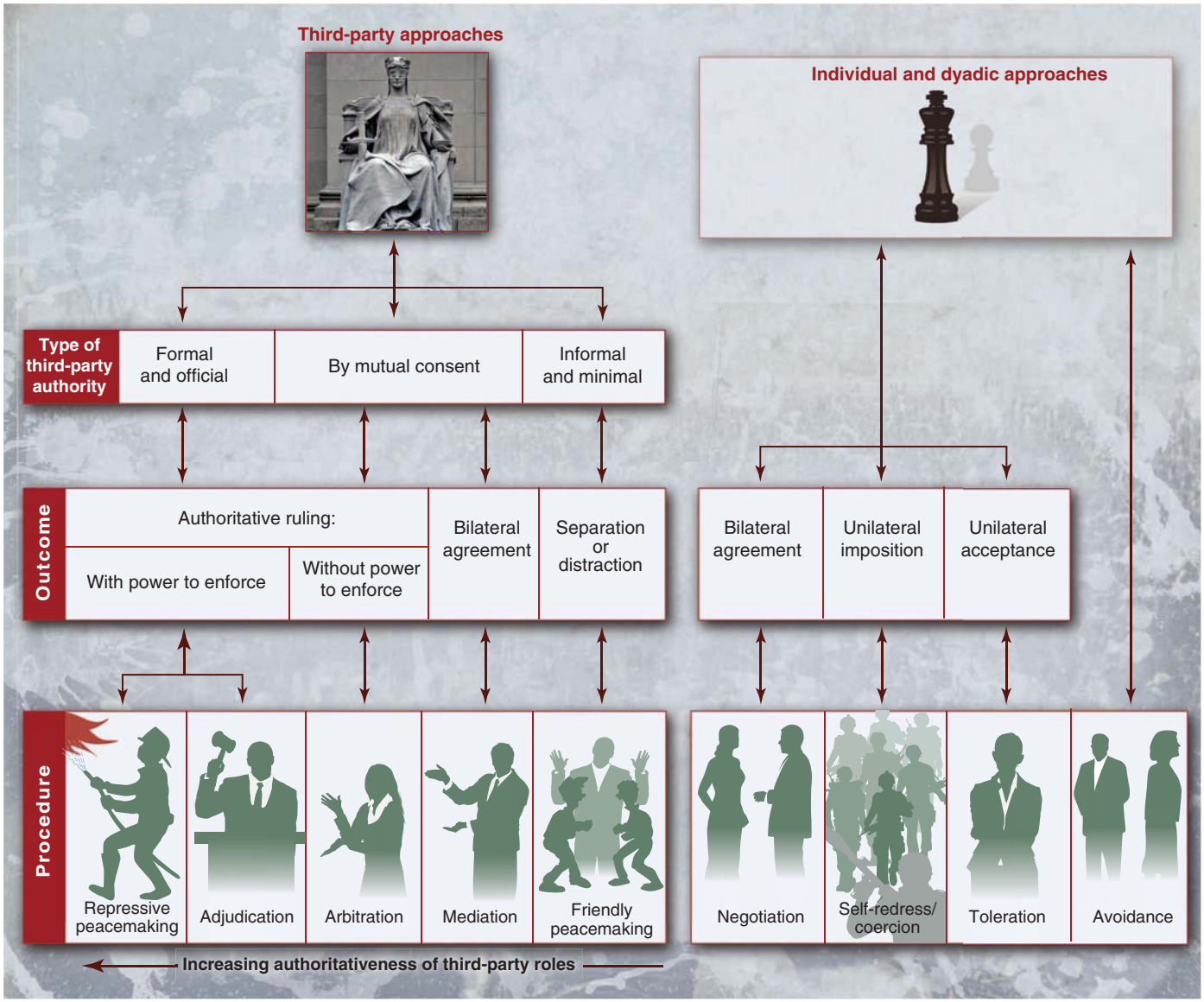


Fig. 2. Approaches to conflict management. Approaches to conflict can entail individual, dyadic, or triadic forms. This typology shows five kinds of third-party assisted procedures (on the left) and four individual or dyad approaches (on the right). Self redress/coercion may involve purely verbal arguments or threats but sometimes entails the use of physical violence (e.g., assault, homicide, feud, or war). Most conflict management approaches do not involve the use of physical violence. Negotiation, for example, is a dyadic procedure wherein parties use techniques like problem-solving and compromising in an attempt to reach an agreement or a resolution of the dispute. In the triadic approaches of adjudication and arbitration, for example, third parties render formal decisions, but adjudication and arbitration differ in that judges have the authority to enforce their rulings, whereas arbitrators do not. Different approaches to conflict management are favored in different cultures, but every society, even where self redress/coercion regularly occurs, has various alternative ways to managing conflict without physical violence (*49*–*51*). [Photo credit: David Dye]

rections (*42*). The eagle perched atop the Tree of Peace was a reminder that one must remain vigilant to threats to the peace. The legends and symbols of peace were regularly recounted at the meetings of the Council of Chiefs (*38*, *42*). Whereas the Iroquois still engaged in external warfare after the confederacy was formed, the main goal was to maintain peace, security, and unity within the confederacy (*39*). They expressed the hope that someday the League of Peace would be extended to include all their neighbors (*38*, *42*).

Superordinate Institutions and Conflict Management

Conflicts within or between groups can be addressed in different ways, most of which do not entail the use of violence, for instance, negotiation, mediation, and adjudication (Fig. 2). Moreover, higher levels of governance can be created with the effect that conflict management among constituent social units becomes more effective and less belligerent (*2*). The creation of the Iroquois Confederacy established a higher level of governance consisting of a Council of Chiefs representing all five nations (and later six when the Tuscarora joined) that assembled to resolve disputes and address other political issues. In terms of structure, the Iroquois employed the same model of village and tribal councils, but scaled up to the supratribal level (*39*). The governing approach was built on discussion and consensus formation. Although there were places on the council for 50 chiefs from major villages across all the nations, each of the nations had only one

vote. The Iroquois core value of peace within their confederacy was personified in the decorum and actions of the Council of Chiefs, who reenacted the legend of how the prophet Deganawidah had originally shown the people of the five nations the path away from feuding and warring among themselves and toward peace, unity, and security under their Tree of Peace (*38*, *42*).

In 1946, Winston Churchill proposed that a pan-European peace could be forged through the creation of strong trade relations and called for the creation of the United States of Europe (*46*, *47*). Likewise, Jean Monnet realized that the centuries of warfare in Europe fundamentally stemmed from a nation-state system and, therefore, to abolish warfare in Europe, a new order with centralized, supranational institutions must be established. A number of leaders, such as Robert Schuman and Konrad Adenauer, adopted the vision of an interdependent and united Europe that would put an end to war in the region once and for all (*45*, *47*). The EU has added a higher level of governance that includes new institutions for dealing with conflict, such as the European Court of Justice, to accomplish the superordinate goals of preventing war and promoting prosperity within the union (*44*, *45*). In both the Iroquois and EU cases, the key was to create a higher-order level of governance, along with a new common identity and a new unity of purpose, to bring about an end to war and to guarantee peace and security within the system. In fact, the same process has taken place in the formation of the United States from 13 original colonies each initially having their own social identities. In the United States today, it is taken for granted that crimes and disputes will be handled though a hierarchy of courts that range from the municipal, state, and district level to the Supreme Court. President Harry Truman once observed, "When Kansas and Colorado have a quarrel over the water in the Arkansas River, they don't call out the National Guard in each state and go to war over it. They bring suit in the Supreme Court of the United States and abide by the decision. There isn't a reason in the world why we cannot do that internationally" (*48*).

Conclusion

Can humanity exist without war? Some completely nonwarring societies and peace systems, such as the tribes of India's Nilgiri and Wynaad plateaus, the Orang Asli societies of mainland Malaysia, and the Australian Aborigines of the Great Western Desert, suggest that this is possible. Furthermore, the Upper Xingu peoples, Iroquois Confederacy, and EU are not mere utopian fantasies; they represent real-world clusters of neighboring societies that live together without war within their peace systems. They have found similar multifaceted paths to successfully keep the peace. These peace systems represent new perspectives and possibilities for living without war that play on the necessity for cooperation under conditions of interdependence.

Could a global peace system be created to abolish war from the planet? Theoretically, this could be done. The people of the earth today face some of the same challenges that Europeans successfully addressed after World War II. Specifically, how can we create security, peace, and prosperity among former enemies and nations with different languages, customs, and cultural traditions? In the 21st century, humanity must determine how to live without war and work together to solve shared challenges like global warming, oceanic pollution, deforestation, desertification, and biodiversity loss that not only threaten particular regions but ultimately endanger human survival overall. When it comes to such threats, all peoples on Earth are interdependent.

Constructing a peace system for the entire planet would involve many synergistic elements, including the transformative vision that a new peace-based global system is in fact possible, the understanding that interdependence and common challenges require cooperation, an added level of social identity that includes all human beings and encompasses more than mere national patriotism, the creation of effective democratic and judicial procedures at a supranational level, and the development of values and symbols that not only sustain peace and justice for all but also relegate the institution of war, like slavery before it, to the pages of history.

References and Notes

1. J. Rubin, D. Pruitt, S. H. Kim, *Social Conflict* (McGraw-Hill, New York, ed. 2, 1994).
2. D. P. Fry, *The Human Potential for Peace* (Oxford Univ. Press, New York, 2006).
3. F. de Waal, *The Age of Empathy* (Harmony Books, New York, 2009).
4. M. A. Nowak, with R. Highfield, *SuperCooperators* (Free Press, New York, 2011).
5. R. W. Sussman, C. R. Cloninger, Eds., *Origins of Altruism and Cooperation* (Springer, New York, 2011).
6. S. Bowles, H. Gintis, *A Cooperative Species* (Princeton Univ. Press, Princeton, NJ, 2011).
7. M. Bekoff, J. Pierce, *Wild Justice* (Univ. of Chicago Press, Chicago, 2009).
8. F. de Waal, *Peacemaking Among Primates* (Harvard Univ. Press, Cambridge, MA, 1989).
9. P. Verbeek, *Behaviour* **145**, 1497 (2008).
10. D. Hart, R. W. Sussman, in *Origins of Altruism and Cooperation*, R. W. Sussman, C. R. Cloninger, Eds. (Springer, New York, 2011), pp. 19–40.
11. K. Weiss, A. Buchanan, *The Mermaid's Tale* (Harvard Univ. Press, Cambridge, MA, 2009).
12. C. Ember, *Ethnology* **17**, 439 (1978).
13. R. Wrangham, D. Peterson, *Demonic Males* (Houghton Mifflin, Boston, 1996).
14. S. Pinker, *The Better Angels of Our Nature* (Viking, New York, 2011).
15. R. B. Ferguson, in *Origins of Altruism and Cooperation*, R. W. Sussman, C. R. Cloninger, Eds. (Springer, New York, 2011), pp. 248–270.
16. T. D. White *et al.*, *Science* **326**, 64 (2009).
17. N. A. Chagnon, *Science* **239**, 985 (1988).
18. S. Beckerman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 8134 (2009).
19. M. Miklikowska, D. P. Fry, in *Nonkilling Psychology*, D. J. Christie, J. E. Pim, Eds. (Center for Global Nonkilling, Honolulu, 2012), pp. 43–70.
20. J. Maynard Smith, G. Price, *Nature* **246**, 15 (1973).
21. D. P. Fry, G. Schober, K. Björkqvist, in *Nonkilling Societies*, J. E. Pim, Ed. (Center for Global Nonkilling, Honolulu, 2010), pp. 101–128.
22. P. Roscoe, *Am. Anthropol.* **109**, 485 (2007).
23. J. B. Silk *et al.*, *Proc. Biol. Sci.* **276**, 3099 (2009).
24. S. Hrdy, *Mothers and Others* (Harvard Univ. Press, Cambridge, MA, 2009).
25. D. P. Fry, in *Origins of Altruism and Cooperation*, R. W. Sussman, C. R. Cloninger, Eds. (Springer, New York, 2011), pp. 227–247.
26. J. Rilling, in *Origins of Altruism and Cooperation*, R. W. Sussman, C. R. Cloninger, Eds. (Springer, New York, 2011), pp. 295–306.
27. D. Grossman, *On Killing* (Little Brown, New York, Rev. ed., 2009).
28. S. Marshall, *Men Against Fire* (Univ. of Oklahoma Press, Norman, 2000).
29. Q. Wright, *A Study of War* (Univ. of Chicago Press, Chicago, 1942).
30. A. Johnson, in *Adaptive Responses of Native Amazonians*, R. Hames, W. Vickers, Eds. (Academic Press, New York, 1983), pp. 61–63.
31. K. Endicott, K. Endicott, *The Headman Was a Woman* (Waveland, Long Grove, IL, 2008).
32. R. Tonkinson, in *Keeping the Peace*, G. Kemp, D. P. Fry, Eds. (Routledge, New York, 2004), pp. 89–104.
33. D. P. Fry, B. Bonta, K. Baszarkiewicz, in *Handbook on Building Cultures of Peace*, Joseph de Rivera, Ed. (Springer, New York, 2009).
34. D. P. Fry, *J. Aggression, Conflict, Peace Res.* **1**, 4 (2009).
35. M. Miklikowska, D. P. Fry, *Beliefs and Values* **2**, 124 (2010).
36. J. Haas, in *Ancient Warfare*, J. Carman, A. Harding, Eds. (Sutton, Gloucestershire, UK, 1999), pp. 11–24.
37. T. Gregor, in *The Anthropology of Peace and Nonviolence*, L. Sponsel, T. Gregor, Eds. (Lynne Rienner, Boulder, CO, 1994).
38. M. Dennis, *Creating a Landscape of Peace* (Cornell Univ. Press, Ithaca, 1993).
39. C. Kupchan, *How Enemies Become Friends* (Princeton Univ. Press, Princeton, NJ, 2010).
40. M. Sherif *et al.*, *The Robbers Cave Experiment: Intergroup Conflict and Cooperation* (Wesleyan Univ. Press, Middletown, CT, 1988).
41. T. Gregor, in *The Anthropology of War*, J. Haas, Ed. (Cambridge Univ. Press, Cambridge, 1990), pp. 105–124.
42. P. Wallace, *White Roots of Peace* (Clear Light, Santa Fe, NM, 1994).
43. J. D. Sachs, *Common Wealth* (Penguin, New York, 2008).
44. A. Staab, *The European Union Explained* (Indiana Univ. Press, Bloomington, IN, 2008).
45. Europa, *Treaty of Lisbon: A Europe of Rights and Values*. Accessed Feb. 15, 2012 at http://europa.eu/lisbon_treaty/glance/rights_values/index_en.htm.
46. S. Hill, *Europe's Promise* (Univ. of California Press, Berkeley, 2010).
47. T. Reid, *The United States of Europe* (Penguin, New York, 2004).
48. T. Hudgens, *We Need Law* (BILC Corporation, Denver, 1986).
49. K.-F. Koch, *War and Peace in Jalemo* (Harvard Univ. Press, Cambridge, MA, 1974).
50. D. Black, *The Social Structure of Right and Wrong* (Academic Press, San Diego, 1993).
51. D. P. Fry, in *Natural Conflict Resolution*, F. Aureli, F. de Waal, Eds. (Univ. of California Press, Berkeley, 2000), pp. 334–351.

Acknowledgments: I thank M. Dennis for graciously sharing his knowledge about the Iroquois Confederacy, my colleagues and students at Åbo Akademi University for fruitful discussions, and three anonymous reviewers for their input. Some of the data reported in this paper were collected during research funded by the National Science Foundation (grant 03-13670).

10.1126/science.1217987

Evolutionary Diversity of the Mitochondrial Calcium Uniporter

Alexander G. Bick,* Sarah E. Calvo,* Vamsi K. Mootha†

Vertebrate mitochondria store large amounts of calcium taken up through a channel called the calcium uniporter. The uniporter, combined with additional intake, buffering, and efflux mechanisms (*1*), is hypothesized to regulate signaling, energy metabolism, and cell death. Although uniporter activity was documented nearly 50 years ago (*2*, *3*), its molecular identity remained elusive until comparative genomics revealed its regulatory partner, MICU1, and pore-forming subunit, MCU (*4*–*6*). MCU represents a fundamentally new class of calcium channels, whereas MICU1 is a peripheral membrane protein with two EF-hand motifs, resembling known calcium-sensing regulators. Molecular identification of the uniporter machinery, combined with the recent availability of diverse genomes, offers an opportunity to explore this channel's evolutionary history.

We examined the distribution of MCU and MICU1 across 138 fully sequenced eukaryotic organisms (*7*). We used sequence similarity followed by inspection of protein domains to annotate organisms that harbor one or more homologs of MCU or of MICU1 and visualized the data on a phylogenomic tree constructed by using standard maximum likelihood methods (*8*).

Homologs of MCU are distributed widely across all major branches of eukaryotic life, present in nearly all plants and metazoa, but with apparent loss events in certain protozoan and fungal lineages (Fig. 1A). MCU homologs have highly conserved residues in the vicinity of a DIME (Asp-Ile-Met-Glu) motif flanked by two transmembrane and coiled-coil domains (Fig. 1B). Within protozoa, MCU homologs are found in organisms from diverse clades, including kinetoplasts (*Trypanosoma cruzi*), heterolobosea (*Naegleria gruberi*), oomycetes (*Phytophthora infestans*), and ciliates (*Tetrahymena thermophila*), but they are absent from other major lineages, including apicomplexa (*Plasmodium falciparum*) and mitosome-containing eukaryotes (*Encephalitozoon cuniculi*, *Giardia lamblia*). MCU homologs are also present in many fungi, including many basidiomycetes and the early-branching *Allomyces macrogynus*, but they are not found in any of 22 yeast fungi such as *Saccharomyces cerevisiae* (*5*). Most MCU-containing species harbor one or two MCU homologs; however, Metaphyta tend to have three to four homologs and, in several interesting cases, contain predicted chloroplast targeting signals (*7*).

Sequenced metazoa, plants, and protozoa typically have homologs of both MCU and MICU1 or lack both proteins (Fig. 1A), but this pattern does not hold in fungi. MCU-containing fungi typically lack MICU1 homologs altogether, suggesting a distinct regulatory mechanism. The two exceptions are deep-branching fungi *A. macrogynus* and *Spizellomyces punctatus*, which suggests that MICU1 was lost subsequent to the divergence of this branch. Fungi also show greater diversity in MCU domain structure, including the circularly permuted arrangements of coiled-coil and transmembrane domains observed in basidiomycetes (Fig. 1B), which perhaps is coincident with the loss of MICU1.

Mammalian MCU and MICU1 physically interact and are strongly coexpressed (*5*). Although *MICU1* and *MCU* are not adjacent to each other in most of the analyzed genomes, in all vertebrates they are adjacent and share a potential bidirectional promoter (Fig. 1C), providing a plausible mechanistic basis for their coordinate expression in vertebrates.

Lastly, we asked whether any bacteria harbor homologs of this new class of calcium channels. Iterative sequence similarity and protein domain searches revealed three diverse bacteria from the Bacteroidetes/Chlorobi group (*Prevotella oris*, *Chlorobium phaeobacteroides*, and *Cytophaga hutchinsonii*) that contain putative MCU homologs that we now term mitochondrial uniporter homologs (*uni*). The *C. hutchinsonii uni* (CHU_0033) is particularly intriguing given its similar domain organization and conservation of key residues essential for calcium transport (*5*, *6*) (Fig. 1B). Functional studies are required to determine whether bacterial *uni* proteins are also selective cationic channels. If shown to transport calcium, they would be among the first identified prokaryotic calcium channels.

Our phylogenomic analysis indicates that the uniporter may have been an early feature of mitochondria. Identification of MCU homologs may help to elucidate the structure of this class of ion channels and the role of mitochondrial calcium transport in the emergence of eukaryotic life.

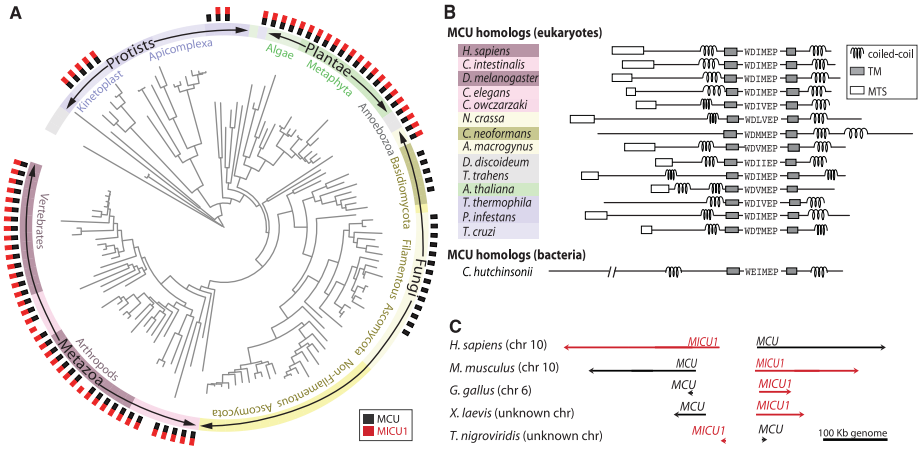


Fig. 1. (**A**) Distribution of *MCU* (black) and *MICU1* (red) homologs across 138 eukaryotes; a detailed version is available in fig. S1. (**B**) Protein domain architecture of selected MCU homologs, including mitochondrial targeting sequence (MTS, white rectangle), coiled-coil domain (coil), transmembrane domain (TM, shaded rectangle), and the DIME motif. L, Leu; T, Thr; P, Pro; V, Val; W, Trp. (**C**) Genomic organization of *MICU1* and *MCU* across vertebrate genomes.

References and Notes

1. G. Hajnóczky, G. Csordás, *Curr. Biol.* **20**, R888 (2010).
2. H. F. Deluca, G. W. Engstrom, *Proc. Natl. Acad. Sci. U.S.A.* **47**, 1744 (1961).
3. F. D. Vasington, J. V. Murphy, *J. Biol. Chem.* **237**, 2670 (1962).
4. F. Perocchi *et al.*, *Nature* **467**, 291 (2010).
5. J. M. Baughman *et al.*, *Nature* **476**, 341 (2011).
6. D. De Stefani, A. Raffaello, E. Teardo, I. Szabò, R. Rizzuto, *Nature* **476**, 336 (2011).
7. Materials and methods are available as supplementary materials on *Science* Online.
8. F. D. Ciccarelli *et al.*, *Science* **311**, 1283 (2006).

Acknowledgments: We thank A. Pringle, K. Zimmerman, E. Koonin, and E. Lander for critical reading of the manuscript. Protein sequences were obtained from the KEGG Organisms Database, Release 58, or the Origins of Multicellularity Sequencing Project at the Broad Institute. This work was supported by NIH funds (R01GM077465 and R01GM097136 to V.K.M. and T32GM007753-33 to A.G.B.). Alignments and tree available at TreeBASE (accession S12416).

Supplementary Materials

www.sciencemag.org/cgi/content/full/336/6083/886/DC1
Materials and Methods
Fig. S1
References (*9*–*29*)

7 October 2011; accepted 28 February 2012
10.1126/science.1214977

Department of Molecular Biology and Department of Medicine, Massachusetts General Hospital, Boston, MA 02114, USA; Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA; Broad Institute, Cambridge, MA 02142, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: vamsi@hms.harvard.edu

Strongly Interacting Rydberg Excitations of a Cold Atomic Gas

Y. O. Dudin and A. Kuzmich*

Highly excited Rydberg atoms have many exaggerated properties. In particular, the interaction strength between such atoms can be varied over an enormous range. In a mesoscopic ensemble, such strong, long-range interactions can be used for fast preparation of desired many-particle states. We generated Rydberg excitations in an ultra-cold atomic gas and subsequently converted them into light. As the principal quantum number n was increased beyond ~70, no more than a single excitation was retrieved from the entire mesoscopic ensemble of atoms. These results hold promise for studies of dynamics and disorder in many-body systems with tunable interactions and for scalable quantum information networks.

A quantum many-particle system in which interactions can be tuned over a vast range may enable profound changes in the way we understand and explore physics of the microscopic realm (*1*). For example, it may lead to previously unknown phases of matter and aid in the discovery of new phenomena. Strong coupling allows for fast, controllable many-body dynamics, whereas the weakly interacting mode can be used for precise external manipulations and measurements. In the context of quantum information science, strong interactions are required to implement fast quantum gates, and long-term storage is achieved in the noninteracting regime (*2*, *3*).

Cold atomic gases are a fruitful platform for such studies. They permit one to perform experiments under well-understood and controlled conditions. Tuning the strength of short-range interactions can be accomplished by using magnetic and optical Feschbach resonances (*1*, *4*). Resonant optical driving of an atomic gas between the ground level and a high-lying Rydberg level is a particularly promising setting for studies of many-body systems with strong, long-range interactions (*5*–*13*). The interaction strength between atoms can vary enormously depending on the principal quantum number n of the atomic level. An important manifestation of Rydberg-level interactions is excitation blockade, in which an atom promoted to a Rydberg level shifts the energy levels of nearby atoms, suppressing their excitation. Previously, excitation blockade has been used to prepare entangled states of two isolated atoms (*14*, *15*) and has also been observed in atomic gases (*16*–*23*). However, controlled creation and manipulation of collective quantum excitations would open the door to investigations of interacting bosonic and fermionic fields, simulations of engineered quantum systems, and generation of complex entangled states of matter and light (*5*–*13*). Encoding of quantum information into symmetrized many-atom quantum states, or spin waves, is particularly attractive because discrimination against various types of noise is built into the manipulation and measurement protocols (*24*–*26*). Atomic spin waves can be efficiently mapped onto light fields, making such systems naturally suited for distribution of quantum information among remote locations.

A small size of the atomic ensemble is a prerequisite for complete suppression of multiply excited spin waves. The ensemble should also have a sufficiently large optical depth to couple the atomic spin wave to the retrieved light field. To create such a small and dense ensemble, we loaded laser-cooled ^{87}Rb atoms into a one-dimensional (1D) optical lattice formed by a retro-reflected (782-nm wavelength) light beam with Gaussian $1/e^2$ waists of $w_y = 15$ μm and $w_z = 55$ μm along the y and z dimensions, respectively (Fig. 1A). The geometry of the lattice and of the excitation beams was chosen so that the ensemble volume is determined by the $w_e = 9$ μm waists of the excitation beams in two dimensions (x and z) and the 15 μm lattice waist in the third (y). Atoms are prepared in the $|5s_{1/2}, F = 2, m_F = 0\rangle$ state by means of optical pumping in a magnetic bias field $B_0 = 4.3$ G applied along the z axis. After preparation, the lattice is turned off in order to avoid differential light shifts, and a 2.5-μs-long experimental cycle is repeated for a period of 200 μs, limited by thermal expansion of the temperature (T) ≅ 10 μK atomic cloud. A τ_e = 200-ns-long period of two-photon excitation by nearly counterpropagating 795 nm (Ω_1) and 475 nm (Ω_2) laser fields tuned near the two-photon $|5s_{1/2}, F = 2\rangle \leftrightarrow |ns_{1/2}\rangle$ resonance creates an optical-frequency spin wave between the ground and the Rydberg levels (Fig. 1B).

After a controlled storage period T_s, a 475-nm retrieval field (Ω_3) resonant with the $|ns\rangle \leftrightarrow |5p_{1/2}\rangle$ transition illuminates atoms for a 1-μs period. This read-out laser pulse converts the spin wave into the retrieved field, which is coupled into a single-mode fiber followed by a beam-splitter and a pair of single-photon detectors, D_1 and D_2. Statistics of the retrieved spin wave is inferred from the distribution of the photoelectric detection events (*27*).

The sum of the photoelectric detection event probabilities at the two detectors, $P = p_1 + p_2$, is shown in Fig. 2A as a function of the two-photon detuning Δ_2 from level $|90s_{1/2}\rangle$. The two peaks correspond to $m = \pm 1/2$ Zeeman sublevels split by the bias field B_0; the width of the peaks, $\gamma \approx 2\pi \times 5$ MHz, is determined by the spectral widths of the excitation laser pulses. In Fig. 2B, the time-resolved photoelectric detection probability P is shown for values of n between 50 and 102. With a $\sim n^{-3/2}$ scaling of the dipole matrix element—and, hence, of the Rabi frequency Ω_3 for

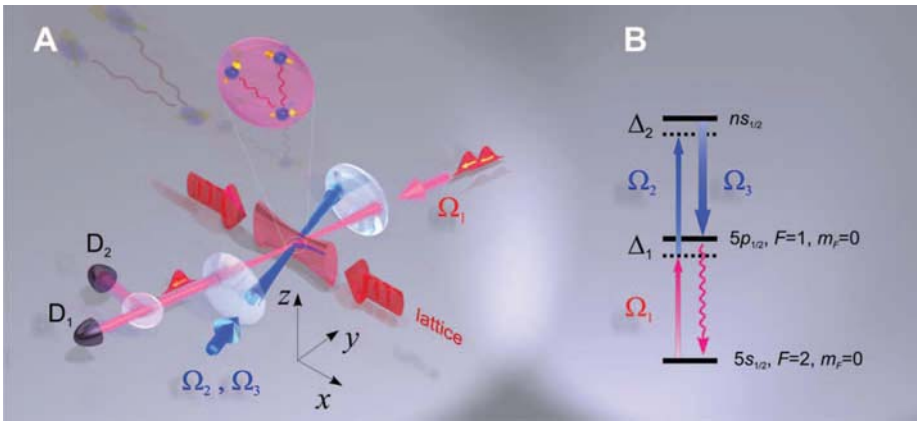


Fig. 1. (**A**) A cold dense sample of atomic ^{87}Rb is prepared in a 1D optical lattice. The lattice is turned off for the experimental sequence, in which nearly counterpropagating 795-nm (Ω_1) and 475-nm (Ω_2) light fields excite a spin wave between the ground $|5s_{1/2}\rangle$ and a Rydberg $|ns_{1/2}\rangle$ level. After a variable delay, a read-out pulse of 475-nm light (Ω_3) converts the Rydberg spin wave into a light field. A Hanbury Brown-Twiss setup of a beamsplitter (BS) followed by two detectors D_1 and D_2 is used to measure the second-order intensity correlation function $g^{(2)}(0)$ of the idler field. (**B**) Relevant ^{87}Rb energy levels. Electronic, hyperfine, and Zeeman quantum numbers are shown. The detuning from the intermediate $|5p_{1/2}\rangle$ level is $\Delta_1 = -40$ MHz; the detuning Δ_2 is varied for the data in Fig. 2A, otherwise it is fixed at the two-photon resonance $|5s_{1/2}, F = 2\rangle \leftrightarrow |ns_{1/2}, m = \pm 1/2\rangle$ between the ground level and one of the Zeeman sublevels of the Rydberg level ($|\Delta_2| \approx 6$ MHz).

School of Physics, Georgia Institute of Technology, Atlanta, GA 30332–0430, USA.

*To whom correspondence should be addressed. E-mail: alex.kuzmich@physics.gatech.edu

the $|5p_{1/2}\rangle \leftrightarrow |ns_{1/2}\rangle$ transition—the retrieved field duration increases with n for a fixed power (18 mW) of the 475-nm read-out field.

In Fig. 2C, P is shown as a function of the two-photon excitation pulse area $\theta \equiv \Omega\tau_e$ for $n = 74$, 81, and 90, where the peak two-photon Rabi frequency is $\Omega \cong (\Omega_1\Omega_2)/(2\Delta_1)$ in the limit $\Omega_{1,2} << \Delta_1$. In the case of strong excitation blockade, collective Rabi oscillations occur even for inhomogeneous atom-light coupling (*28*). Technical sources of dephasing, such as laser linewidth, ac Stark shifts, and atom number fluctuations, are expected to reduce P to half of its maximum value as θ is increased. In contrast, in Fig. 2C we observe a decay of P to 0, with no revivals. These observations are consistent with the dephasing of multiply excited spin waves (*13*). The data for $n = 90$ are fit with a function $\zeta\mathcal{P}\theta^2\exp(-\mathcal{P}\theta^2)$, where $\zeta = 0.060(1)$ and $\mathcal{P} = 51(1)$ are adjustable parameters. The choice of the fit function is suggested by a model with Poisson spin-wave excitation statistics (*13*, *29*). Within the model, ζ corresponds to the overall measured retrieval and detection efficiency for the spin-wave excitation, whereas the maximum photoelectric detection probability is $P_m = \zeta/e$.

It is instructive to compare the strongly interacting regime to the ideal limit of a noninteracting ensemble, in which each atom undergoes a Rabi oscillation between the ground level $|5s_{1/2}\rangle$ and the Rydberg level $|ns_{1/2}\rangle$ with the position-dependent two-photon Rabi frequency $\Omega(\mathbf{r})$. Atoms in a sufficiently low-n Rydberg level may be used to approximate this ideal situation. In Fig. 2D, $P/\mathcal{T}$ is displayed as a function of θ both for high- and low-n Rydberg levels. The retrieved field is attenuated (for low-n) by a factor $1/\mathcal{T}$ to prevent saturation of the single-photon detectors D_1 and D_2. The weakly interacting (low-n) regime is represented by measurements with levels $|19s_{1/2}\rangle$ and $|21s_{1/2}\rangle$ ($\mathcal{T}_{19,21} = 0.05$). The maximum photoelectric detection probability P_m and the corresponding pulse area θ_m are lower for $n = 21$ than for $n = 19$, which is indicative of interaction-induced excitation suppression. Results for $n = 74$, 81, and 90 ($\mathcal{T}_{74,81,90} = 1$) suggest that for such high n, both P_m and θ_m approach asymptotic values. This would be expected if only a single retrievable excitation is generated in the entire ensemble. In this picture, the effective number of atoms $\mathcal{N}$ in the ensemble is proportional to the ratio of $P_m/\mathcal{T}$ for $n = 19$ and $n = 90$, giving $\mathcal{N} \cong 5 \times 10^2$. Collectively, enhanced coupling of the driving laser fields to the singly excited spin wave implies $\mathcal{N} \sim [\theta_m\,(n = 19)/\theta_m(n = 90)]^2 \cong 6 \times 10^2$. In the absence of data for $n < 19$, these values should be considered as low-end estimates of $\mathcal{N}$.

The single-photon character of the retrieved field was explicitly confirmed by cross-correlating the photoelectric counting events at detectors D_1 and D_2 as a function of time delay, as shown for upper-level $|102s_{1/2}\rangle$ in Fig. 3, inset. The observed coincidences around zero time delay are strongly suppressed, providing evidence for the quantum nature of the emitted light. We further investigated the transition to the regime of strongly interacting spin waves by measuring the second-order intensity correlation function at zero delay $g^{(2)}(0)$ (*27*) as a function of effective principal quantum number $n^* = n - \delta_s$ (Fig. 3). Here, $\delta_s = 3.13$ is the s-wave quantum defect for Rb. As n^* is increased, interaction-induced suppression of spin waves with more than one Rydberg atom sets in, with the transition to the single retrievable excitation regime occurring for $n \sim 60$ to 70. The transition is associated with the interaction strength scaling as n^4/R^3 in the dipole-dipole regime for $R \lesssim R_c$ and as n^{11}/R^6 in the van der Waals regime for $R \gtrsim R_c$. Here, R_c is the critical inter-atom distance (*2*). For $n \lesssim 60$, van der Waals interactions are

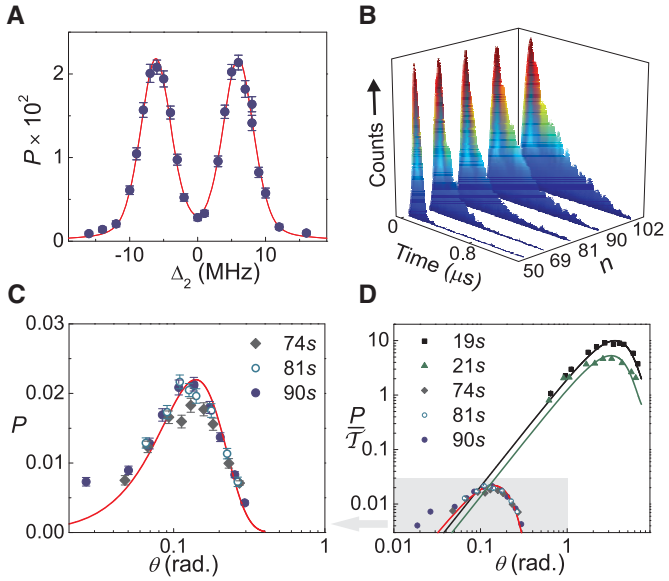


Fig. 2. (**A**) Probability P of photoelectric detection event per trial as a function of two-photon detuning Δ_2 shows the $m = \pm 1/2$ Zeeman components of level $|90s_{1/2}\rangle$ split by the bias field B_0. The solid curve is a pseudo-Voigt fit. (**B**) Normalized temporal profiles of the retrieved field for upper levels $|ns_{1/2}\rangle$ for n between 50 and 102, with a fixed power of the retrieval field Ω_3. (**C**) P as a function of (single-atom) two-photon excitation pulse area θ for $T_s = 0.4$ μs: solid circles for $|90s_{1/2}\rangle$, open circles for $|81s_{1/2}\rangle$, and diamonds for $|74s_{1/2}\rangle$ upper level. The solid curve is a fit of the form $\zeta\mathcal{P}\theta^2\exp(-\mathcal{P}\theta^2)$ to the $|90s_{1/2}\rangle$ data. (**D**) $P/\mathcal{T}$ as a function of θ. Results plotted in (C) are shown together with additional data for levels $|19s_{1/2}\rangle$ and $|21s_{1/2}\rangle$. The data for $n = 19$ and $n = 21$ are taken with the retrieved field attenuated by a factor $1/\mathcal{T}_{19,20} = 20$ to avoid detector saturation, whereas for $n = 74$, 81, and 90, no attenuation is used ($\mathcal{T}_{74,81,90} = 1$). The $n = 19$ and $n = 21$ data are fitted with a function accounting for averaging of sinusoidal oscillations of the retrievable spin-wave amplitude across the Gaussian transverse spatial profiles of $\Omega_{1,2}(\mathbf{r})$ (*27*). Error bars represent ±1 SD ($\sqrt{M}$) for M photoelectric counting events.

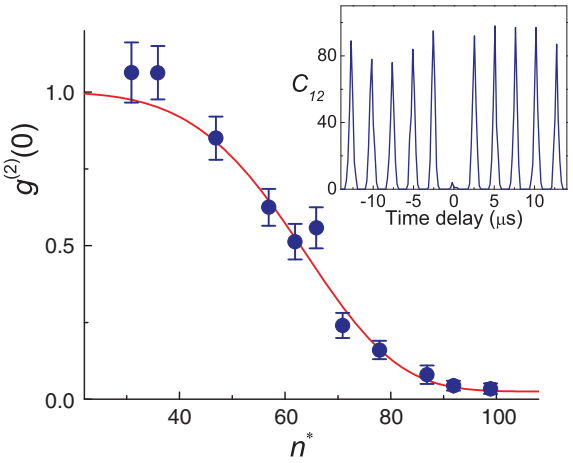


Fig. 3. Measured second-order intensity correlation function at zero time delay $g^{(2)}(0)$ as a function of the effective principal quantum number $n^* = n - \delta_s$ of the upper-level $|ns_{1/2}\rangle$ for $T_s \cong 0.3$ μs. For $n^* > 70$, the data are taken at θ_m, whereas for $n^* \lesssim 70$, the values of θ are chosen to keep P between 0.02 and 0.03 (Fig. 2D). The solid line is a fit of the form $g^{(2)}(0) = [1 - g^{(2)}_{bg}]\exp[-(n^*/n^*_0)^\alpha] + g^{(2)}_{bg}$, with best-fit values $\alpha = 4.7(4)$ and $n^*_0 = 67(1)$. (**Inset**) Cross-correlated coincidence counts C_{12} as a function of time delay for upper-level $|102s_{1/2}\rangle$. Error bars represent ±1 SD ($\sqrt{M}$) for M photoelectric counting events.

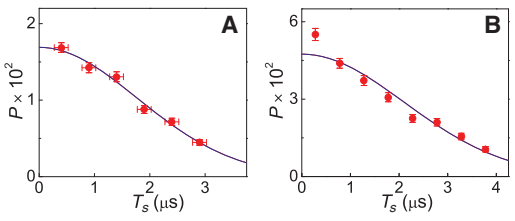


Fig. 4. Photoelectric detection probability P as a function of the storage time T_s. (**A**) For upper-level $|90s_{1/2}\rangle$. (**B**) For upper-level $|34s_{1/2}\rangle$. Experimental data showing decay are fitted with a Gaussian function $\exp(-T_s^2/\tau^2)$, with best-fit values $\tau_{90} = 2.5(1)$ μs and $\tau_{34} = 3.0(1)$ μs. The vertical error bars represent ±1 SD ($\sqrt{M}$) for M photoelectric counting events. The horizontal error bars represent the lengths of the retrieved light pulses.

relevant because $R_c \lesssim 3$ μm is sufficiently smaller than the ~10-μm sample size. In contrast, for $n \sim 100$ dipole-dipole interactions become operational as $R_c \cong 10$ μm.

High-quality single-photon statistics are observed for $n \gtrsim 90$. For an ideal single-photon source, $g^2(0)$ would be nonzero only because of background counts on $D_{1,2}$. The measured background photoelectric count probability is $P_B = 2.5 \times 10^{-4}$ for the 0.74-μs detection window, with $P_D = 0.55P_B$ resulting from detector dark counts, and the remainder caused by various scattered light sources. The measured value of $g^{(2)}(0) = 0.040(14)$ for level $|102s_{1/2}\rangle$ is consistent with a lower bound of $g^{(2)}_{bg} = 0.025(1)$ because of background counts. A dearth of excess coincidences indicates a strong suppression of spin waves with more than one excitation. Both excitation blockade and spin-wave dephasing can contribute to suppression of two-photon events. As a step to separating the roles of the two mechanisms, we used $|nd_{3/2}\rangle$ upper levels for which the excitation blockade is expected to break down because of a pronounced angular dependence of interactions between atom pairs (*2*). The observed strong suppression of two-photon events [$g^{(2)}(0) = 0.066(38)$ for level $|88d_{3/2}\rangle$ and $g^{(2)}(0) = 0.075(26)$ for level $|100d_{3/2}\rangle$] suggests a substantial role of spin-wave dephasing in our experiment.

The measured probability of photoelectric detection per trial P was 2 to 3% for the data in Fig. 3. When normalized by the 0.52 measured optical transmission from the sample to the detectors and 0.55 single-photon detector efficiency, $P_m(n = 90) = 2.8\%$ corresponds to $\epsilon \approx 0.10$ single-photon generation efficiency. This is limited by several experimental imperfections. The retrieval field Ω_3 has the same spatial mode as the excitation field Ω_2. Higher available laser power and a larger spatial mode for Ω_3 are expected to increase ϵ. A smaller sample should lead to a pronounced excitation blockade, whereas a medium-to-low finesse cavity may increase spin wave retrieval efficiency. As it is, the single-photon source has lower $g^{(2)}(0)$ and higher P and is more than three orders of magnitude faster than a deterministic single-photon source (*30*) based on the Duan, Lukin, Cirac, and Zoller protocol (*24*).

To evaluate coherence properties of the created optical spin-wave excitations, the retrieved signal is measured as a function of storage time T_s. For this data, an extended 5-μs-long timing sequence was used. For $n = 90$, the signal is shown in Fig. 4A. The observed $1/e$ coherence time $\tau_{90} = 2.5(1)$ μs is consistent with the decoherence expected from atomic motion, which smears the phase grating imprinted on the atoms during the excitation. The period of the spin wave formed by counterpropagating Ω_1 and Ω_2 fields is $\Lambda \cong 1$ μm. The expected motional coherence time is $\tau \cong \Lambda/(2\pi v) \cong 5$ μs, where $v = (k_B T/M)^{1/2} \cong 3$ cm/s is the atomic velocity, k_B is Boltzmann's constant, $T = 10$ μK is the atomic temperature, and M is the mass of ^{87}Rb. In principle, there may be additional sources of decoherence for highly excited Rydberg levels, such as stray electric fields. To test this, a similar measurement was performed for $n = 34$ (Fig. 4B). The excitation level was kept low to preclude influence of interactions of multiply excited spin waves. The somewhat higher value of the observed coherence time $\tau_{34} = 3.0(1)$ μs may indicate a possible contribution to τ_{90} of Rydberg atom interactions with other atoms or external electric fields. However, atomic motion appears to provide the dominant contribution to the observed decoherence, which is also confirmed by lower values of τ in measurements with higher sample temperatures. In the future, it should be possible to reduce motional dephasing by further cooling. One may also use multi-photon excitation to produce larger-period spin waves, which are less susceptible to atomic motion (*13*), or pin the spin wave in a magic-wavelength (for the ground-Rydberg transition) optical lattice (*31*).

We have shown that strong Rydberg-level interactions can be used for the fast preparation of single-quantum excitations of a cold atomic gas. Using two different spin waves within the same ensemble, each driven by an independent single-photon field Ω_1, a photon-photon quantum gate can be realized. This work can be extended to the creation of several entangled spin waves (*13*). This would also allow investigation of interacting fermionic and bosonic collective excitations (*8*). Entangled spin waves may also be mapped either onto long-lived atomic coherences or light fields, opening possibilities for scalable remote entanglement distribution (*10*, *13*).

References and Notes

1. I. Bloch, J. Dalibard, W. Zwerger, *Rev. Mod. Phys.* **80**, 885 (2008).
2. M. Saffman, T. G. Walker, K. Mølmer, *Rev. Mod. Phys.* **82**, 2313 (2010).
3. A. G. Radnaev *et al.*, *Nat. Phys.* **6**, 894 (2010).
4. M. W. Zwierlein, J. R. Abo-Shaeer, A. Schirotzek, C. H. Schunck, W. Ketterle, *Nature* **435**, 1047 (2005).
5. M. D. Lukin, M. Fleischhauer, R. Cote, *Phys. Rev. Lett.* **87**, 037901 (2001).
6. M. Saffman, T. G. Walker, *Phys. Rev. A* **72**, 042302 (2005).
7. M. Müller, I. Lesanovsky, H. Weimer, H. P. Büchler, P. Zoller, *Phys. Rev. Lett.* **102**, 170502 (2009).
8. B. Olmos, R. González-Férez, I. Lesanovsky, *Phys. Rev. Lett.* **103**, 185302 (2009).
9. H. Weimer, M. Müller, I. Lesanovsky, P. Zoller, H. P. Büchler, *Nat. Phys.* **6**, 382 (2010).
10. B. Zhao, M. Müller, K. Hammerer, P. Zoller, *Phys. Rev. A* **81**, 052329 (2010).
11. G. Pupillo, A. Micheli, M. Boninsegni, I. Lesanovsky, P. Zoller, *Phys. Rev. Lett.* **104**, 223002 (2010).
12. A. V. Gorshkov, J. Otterbach, M. Fleischhauer, T. Pohl, M. D. Lukin, *Phys. Rev. Lett.* **107**, 133602 (2011).
13. F. Bariani, Y. O. Dudin, T. A. B. Kennedy, A. Kuzmich, *Phys. Rev. Lett.* **108**, 030501 (2012).
14. E. Urban *et al.*, *Nat. Phys.* **5**, 110 (2009).
15. A. Gaëtan *et al.*, *Nat. Phys.* **5**, 115 (2009).
16. K. Singer, M. Reetz-Lamour, T. Amthor, L. G. Marcassa, M. Weidemüller, *Phys. Rev. Lett.* **93**, 163001 (2004).
17. D. Tong *et al.*, *Phys. Rev. Lett.* **93**, 063001 (2004).
18. T. C. Liebisch, A. Reinhard, P. R. Berman, G. Raithel, *Phys. Rev. Lett.* **95**, 253002 (2005).
19. T. Vogt *et al.*, *Phys. Rev. Lett.* **97**, 083003 (2006).
20. R. Heidemann *et al.*, *Phys. Rev. Lett.* **99**, 163601 (2007).
21. E. Brekke, J. O. Day, T. G. Walker, *Phys. Rev. A* **78**, 063830 (2008).
22. H. Kübler, J. P. Shaffer, T. Baluktsian, R. Löw, T. Pfau, *Nat. Photonics* **4**, 112 (2010).
23. J. D. Pritchard *et al.*, *Phys. Rev. Lett.* **105**, 193603 (2010).
24. L.-M. Duan, M. D. Lukin, J. I. Cirac, P. Zoller, *Nature* **414**, 413 (2001).
25. D. N. Matsukevich, A. Kuzmich, *Science* **306**, 663 (2004).
26. E. Brion, K. Mølmer, M. Saffman, *Phys. Rev. Lett.* **99**, 260501 (2007).
27. Supplementary materials are available on *Science* Online.
28. F. W. Cummings, A. Dorri, *Phys. Rev. A* **28**, 2282 (1983).
29. F. Bariani, T. A. B. Kennedy, *Phys. Rev. A* **85**, 033811 (2012).
30. R. Zhao *et al.*, *Nat. Phys.* **5**, 100 (2009).
31. S. Zhang, F. Robicheaux, M. Saffman, *Phys. Rev. A* **84**, 043408 (2011).

Acknowledgments: We thank F. Bariani, B. Kennedy, A. Radnaev, A. Zangwill, and C. Campbell for discussions. This work was supported by the Air Force Office of Scientific Research Atomic Physics Program, Quantum Memories Multidisciplinary University Research Initiative, and NSF.

Supplementary Materials

www.sciencemag.org/cgi/content/full/science.1217901/DC1
Supplementary Text
Fig. S1

14 December 2011; accepted 9 April 2012
Published online 19 April 2012;
10.1126/science.1217901

Water-Mediated Proton Hopping on an Iron Oxide Surface

Lindsay R. Merte,[1]* Guowen Peng,[2] Ralf Bechstein,[1] Felix Rieboldt,[1] Carrie A. Farberow,[2] Lars C. Grabow,[2]† Wilhelmine Kudernatsch,[1] Stefan Wendt,[1] Erik Lægsgaard,[1] Manos Mavrikakis,[2]‡ Flemming Besenbacher[1]‡

The diffusion of hydrogen atoms across solid oxide surfaces is often assumed to be accelerated by the presence of water molecules. Here we present a high-resolution, high-speed scanning tunneling microscopy (STM) study of the diffusion of H atoms on an FeO thin film. STM movies directly reveal a water-mediated hydrogen diffusion mechanism on the oxide surface at temperatures between 100 and 300 kelvin. Density functional theory calculations and isotope-exchange experiments confirm the STM observations, and a proton-transfer mechanism that proceeds via an H_3O^+-like transition state is revealed. This mechanism differs from that observed previously for rutile TiO_2(110), where water dissociation is a key step in proton diffusion.

The diffusion of hydrogen on oxide surfaces is an important process in a number of applications, including catalytic hydrogen evolution and reforming (*1*–*3*), photocatalytic dehydrogenation (*4*), the synthesis of metallic materials from oxide precursors (*5*),

and hydrogen storage (*6*). Diffusion of hydrogen is particularly important in cases where "spillover" effects operate, where the active sites for H_2 dissociation or association are spatially separated from the active sites for other reactions (*1*). For example, on many hydrogenation catalysts, which consist of metal nanoparticles dispersed onto a high–surface area oxide support, H atoms produced by dissociation at the metal particles can migrate from the metal particles onto the oxide and across the oxide surface to another adsorbed reactant. Mechanisms involving spillover hydrogen have been implicated in many important catalytic reactions, including methanol synthesis (*7*) and hydrogenation of benzene and toluene (*8*). The presence of spillover hydrogen on oxides can also affect the catalytic activity of the oxides themselves, in some cases making an otherwise inactive catalyst active (*1*).

Enhanced rates of surface reactions involving hydrogen diffusion upon addition of water have been observed in a number of cases (*4*, *5*, *7*, *9*–*11*). However, the fundamental atomistic mechanisms underlying hydrogen diffusion on solid surfaces and the factors influencing them still remain unsettled because of the great complexity of oxide materials and the difficulty in characterizing them at the atomic level. The use of atomically flat single-crystal surfaces as model systems for more complex materials allows detailed, atomic-scale characterization of surface reactions under well-controlled conditions, and this so-called surface science approach has proven extremely powerful in unraveling the complex processes occurring on the surfaces of heterogeneous catalysts and related materials (*12*). Scanning probe microscopies are capable of revealing the dynamic structure of materials at the atomic scale and can thus provide direct insight into reaction mechanisms and kinetics. However, experimental studies of hydrogen diffusion on well-defined model oxide surfaces are scarce (*13*–*15*) and have been limited almost exclusively to rutile TiO_2(110).

Here, we investigated proton hopping on an oxygen-terminated FeO(111) monolayer film on Pt(111). The FeO monolayer thin film, which has been extensively characterized by a variety of experimental techniques, exhibits a simple hexagonal structure, with a characteristic ($\sqrt{91} \times \sqrt{91}$)R5.2° moiré-type coincidence superlattice structure (*16*) caused by the mismatch between the FeO and the Pt(111) hexagonal layers. An atomically resolved STM image of the bare FeO layer (Fig. 1A) shows individual Fe and O atoms as well as a longer-scale (~2.5 nm) modulation in brightness from the moiré structure. Previous temperature-programmed desorption (TPD) and infrared reflection absorption spectroscopy (IRAS) studies have shown that water molecules adsorb weakly on this surface (*17*–*19*), desorbing at temperatures comparable to water multilayers and ~80 to 100 K lower than on TiO_2(110) and MgO(100), where interaction with exposed metal cations dominates the adsorption behavior of water (*20*). On thicker Fe_3O_4(111) films, the availability of reactive Fe sites on the Fe_3O_4 surface causes water to adsorb dissociatively (*18*). Thus, the Fe atoms of the FeO film must interact very weakly with adsorbed H_2O molecules, so OH-H_2O interactions can be studied in the absence of water dissociation. Note that on TiO_2(110), water dissociation is critical for water-assisted proton hopping. Additionally, by using an atomic hydrogen source, H adatoms bound to FeO lattice oxygen can be produced in a well-controlled manner, as discussed previously (*21*, *22*).

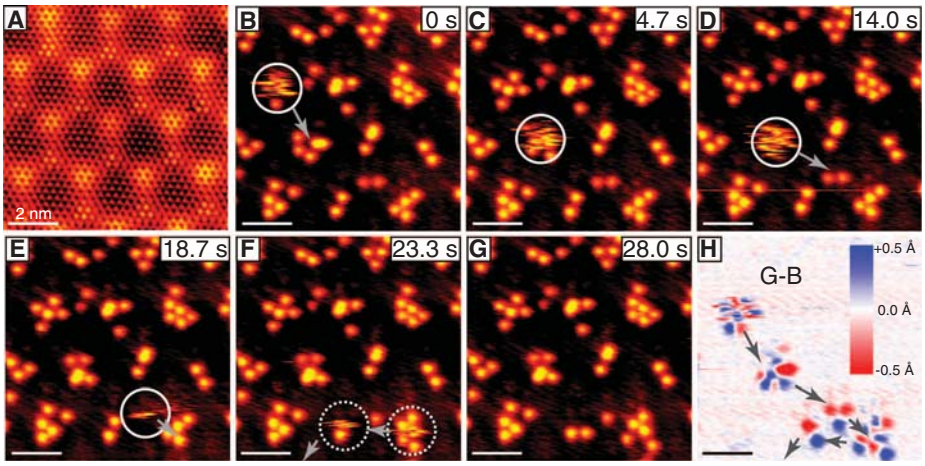


Fig. 1. (**A**) An atomically resolved STM image of the bare FeO film, showing its moiré structure and protrusions due to individual Fe and O atoms. (**B** to **G**) Series of STM images of the hydrogenated FeO film on Pt(111) at 105 K. The hydrogen atoms (bright spots) diffuse rapidly in the presence of a water molecule that hops from one domain to the next until it leaves the scanning area. The rapid movement of H atoms in the presence of the water molecule manifests itself as irregular streaks in the STM image, marked with circles. (**H**) The difference image, obtained by subtracting the first image from the last, shows the changes in H atom positions along the path of the water molecule. All scale bars, 2 nm.

Figure 1, B to G, depicts a series of STM images of the hydrogenated FeO film, where bright protrusions correspond to single H adatoms. The H atoms preferred certain sites in the film's moiré superstructure, which caused them to organize in distinct groups. The temporal sequence of STM images of the hydrogenated film, obtained at 105 K, shows the movements of H atoms induced by a water molecule that moved quickly across the surface. Favorable bonding interactions caused the water molecule to be momentarily trapped at a group of surface-bound H atoms before moving to another nearby group. While the water molecule was trapped at such a cluster, the H atoms within the group rearranged rapidly and repeatedly on the time scale of the STM measurement (~18 ms per scan line), as revealed by the irregular streaked appearance of the moving H atoms in the STM images. After the water molecule moved away, the H atoms changed positions. The difference image in Fig. 1H shows clearly the changes in positions of the H atoms during the course of the STM measurements, and thus the path taken by the water molecule as it crossed this area of the surface.

Water-mediated H atom diffusion was observed up to 300 K, but at this highest sample temperature, the diffusion events were too rapid to follow the motion of the individual mediating water species by STM, as was possible at 105 K. However, the underlying mediated diffusion mechanism (as opposed to the "intrinsic" direct hopping of an H atom from one site to another) was nonetheless reflected in the statistical characteristics of the diffusion rate. Figure 2A shows a series of consecutively acquired STM difference images of the 180 Å × 180 Å area of the surface depicted in the leftmost panel (see movie S1). As in Fig. 1H, the movements of the H atoms between STM scans are visible in the difference image as pairs of red and blue spots. What is striking about the room-temperature measurements is that the number of hopping events occurring in each scan varies enormously from one image to the next. If the hopping events were independent of one another, a more stable hopping rate would be observed, with variability following a Poisson distribution. In fact, the majority of STM images show basically no movement of H atoms, but in a few images, almost all H atoms moved. Figure 2B quantifies the hopping rates from a room-temperature STM movie (movie S2) acquired at a rate of 6 s per scan; the corresponding histogram of scan-to-scan hopping rates is shown in Fig. 2C. The

[1]Interdisciplinary Nanoscience Center (iNANO) and Department of Physics and Astronomy, Aarhus University, DK-8000 Aarhus C, Denmark. [2]Department of Chemical and Biological Engineering, University of Wisconsin, Madison, WI 53706, USA.

*Present address: Department of Physics, University of Central Florida, Orlando, FL 32816, USA.
†Present address: Department of Chemical and Biomolecular Engineering, University of Houston, Houston, TX 77204, USA.
‡To whom correspondence should be addressed. E-mail: fbe@inano.au.dk (F.B.); manos@engr.wisc.edu (M.M.)

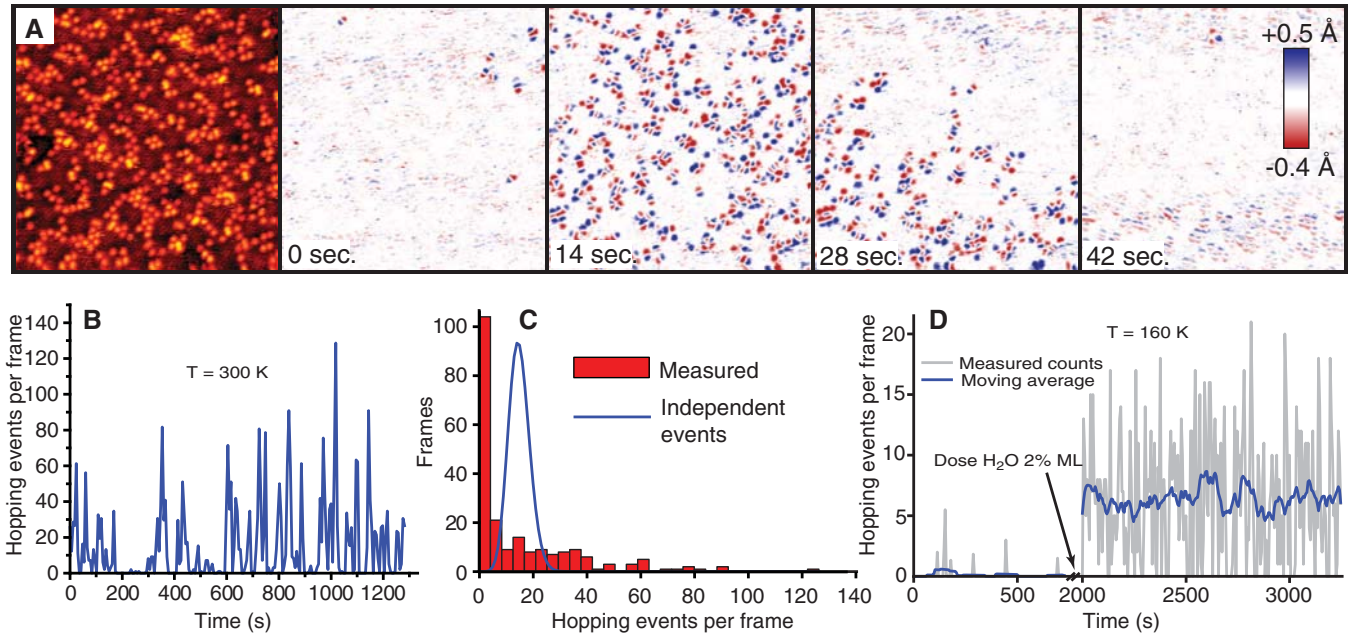


Fig. 2. (**A**) Time-resolved STM measurements of hydrogen atoms on FeO(111)/Pt(111) at room temperature. The leftmost frame shows a single 180 Å × 180 Å STM image of the FeO film; H atoms are visible as bright protrusions. The remaining frames are sequential difference images showing the changes in H atom positions between frames. (**B**) The hopping rate extracted from a room-temperature STM movie, in terms of the number of hopping events observed per frame. (**C**) Histogram of hopping events from (B) compared with that expected for independent diffusion events. (**D**) Hopping rate extracted from STM movies acquired at 160 K before and after exposure to a small amount of water.

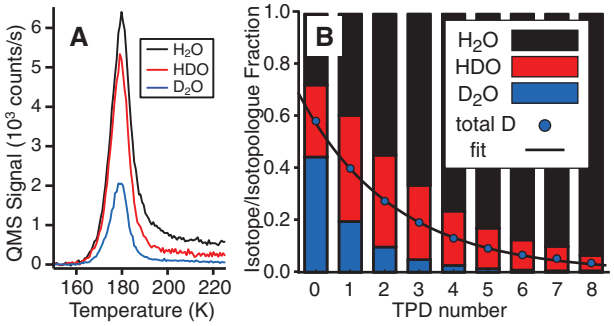


Fig. 3. Isotope exchange between adsorbed D atoms and dosed H_2O. (**A**) Example of a TPD measurement where H_2O was dosed onto a D-covered surface. The peaks at 180 K due to HDO (mass/charge ratio m/z = 19) and D_2O (m/z = 20), in addition to that of H_2O (m/z = 18), arise from proton exchange with the surface. (**B**) Fraction of each isotopologue and of total deuterium in the desorbing water, from a series of TPD measurements where the sample was recooled after each measurement and exposed again to H_2O. Formation of HDO and D_2O depletes the deuterium fraction, which decays exponentially with the number of measurements.

average hopping rate observed in these measurements was 14.9 events per frame, and the deviation from the plot expected for a Poisson distribution for independent events occurring at this rate is apparent. The hopping events are correlated in time, and larger-scale STM measurements further reveal correlations in space on the scale of tens of nanometers (fig. S1) (*23*).

These correlations are easily explained by a diffusion mechanism where mediating species diffuse across the surface and induce diffusion of nearby H atoms. Even under ultrahigh-vacuum (UHV) conditions (~1×10^{-10} to 5×10^{-10} mbar), the number of water molecules impinging on the sample surface from residual gas in the UHV chamber was sufficient to dominate the mobility of the H atoms. Experiments where water molecules were intentionally dosed into the chamber while recording STM movies show a clear correlation between the water pressure and the rate of H atom diffusion (fig. S2). In STM movies recorded at 160 K, a large increase in the hopping rate occurred after dosing a small amount of water onto the surface, as demonstrated in Fig. 2D (see movies S3 and S4, recorded before and after water dosing, respectively).

The acceleration of hydrogen diffusion by water is generally attributed to proton transfer processes, where the water molecule donates one of its protons to a surface site and accepts a proton from another site (*2*). To determine whether proton exchange processes played a role on the observed water-promoted hydrogen diffusion on FeO, we conducted a series of isotope exchange experiments where atomic deuterium (D) was used instead of atomic hydrogen to produce a hydrogenated FeO film (Fig. 3). These experiments were performed by repeatedly dosing H_2O at 130 K to such a deuterated FeO thin film and measuring the TPD yields of H_2O, HDO, and D_2O. Each TPD experiment was stopped at 250 K to avoid reduction of the FeO film (*22*) and to minimize exposure of the sample to the background gas above the water desorption temperature. Desorption of water from the sample surface was observed at ~180 K; sharp peaks are seen for all three isotopologues (Fig. 3A). The substantial yields of HDO and D_2O reflect proton exchange between the water molecules and the deuterated surface. Repeated H_2O doses led to an exponential decay of the deuterium fraction (Fig. 3B), and the rate of this decay (determined by the relative hydrogen and water coverages; fig. S3) indicates that complete isotopic scrambling, and thus facile proton exchange, occurred at temperatures below 160 K.

To elucidate the underlying mechanism of the fast hydrogen diffusion observed in experiments, we performed density functional theory corrected for on-site Coulomb interactions (DFT+U) calculations using the VASP code (*24*) with the generalized gradient approximation (*25*) and projector-augmented wave potentials (*26*, *27*). To mimic the conditions of the experiments as closely as possible, we modeled the FeO/Pt(111) system with the natural $(\sqrt{91} \times \sqrt{91})$R5.2° unit cell and studied the water-mediated H diffusion on the face-centered cubic (fcc) domain, which is the preferred domain for H adsorption (*23*). On the fcc domain of the clean FeO/Pt(111) surface,

water is weakly bound with a calculated binding energy of only –0.11 eV. However, in the presence of a pre-adsorbed H, the binding energy of a water molecule is –0.60 eV. This enhanced binding of a water molecule is caused by the formation of a hydrogen bond with the pre-adsorbed surface H. For H diffusion to occur at temperatures as low as 105 K, the most likely mechanism is a proton transfer process mediated by H_2O (Fig. 4A, insets). The reaction coordinate corresponding to the minimum-energy pathway (blue solid line in Fig. 4A) involves the transfer of the surface proton to the water molecule and the formation of a short-lived hydronium ion (H_3O^+) species at the transition state. Subsequently, one of the water protons is transferred to an O atom on the surface. For this process, we calculated an activation energy barrier of 0.21 eV. This activation energy barrier was found to be invariant with respect to H coverage up to a local H coverage similar to that in the experiments.

For comparison, we also considered the intrinsic diffusion of H atoms in the absence of H_2O molecules. In particular, we studied the direct hopping of a single H atom from one O atom to a neighboring O in the same fcc domain. This direct proton hopping is characterized by a large activation energy barrier, ~1.02 eV (red dashed line in Fig. 4A), which is comparable to that for intrinsic H diffusion on TiO_2(110) (*14*). Thus, for temperatures below room temperature, this pathway can be excluded with certainty.

For water molecules to promote H diffusion as observed experimentally, low activation energy barriers for water diffusion are required, in addition to proton exchange steps. To ensure that this is indeed the case, we used DFT+U calculations to study the diffusion of water in the presence of surface hydroxyls arranged in a $(\sqrt{3} \times \sqrt{3})$R30° overlayer. More specifically, we analyzed water rotation around a surface OH group and water hopping from one surface OH group to a neighboring one. The calculated minimum-energy profile for these two processes is shown in Fig. 4B. The water molecule, adsorbed near a surface OH group, can rotate around this OH group (i.e., in the form of a H_2O-H complex) with small barriers of 0.07 to 0.19 eV (structure i → structure ii → structure iii). Direct transfer of the H_2O molecule from one surface OH to another OH via a flip of the H_2O molecule (structure i → structure iv) is also very facile, with an energy barrier of 0.26 eV. The combination of these steps allows the water molecule to change positions rapidly within a group of adsorbed H atoms, as observed in the STM images shown in Fig. 1.

Water-mediated diffusion of H atoms on the FeO(111) thin film occurs even though no Lewis acid sites (i.e., undercoordinated metal ions) exist on this surface that can bind the water molecules strongly. Thus, the water-mediated H diffusion on FeO(111) is fundamentally different from that previously observed for water-mediated H diffusion on rutile TiO_2(110). On the anisotropic rutile TiO_2(110) surface, across-row proton hopping was found to be facilitated by dissociation and reforming of H_2O molecules at undercoordinated surface Ti^{4+} sites (*15*). This process is only possible because of a particular arrangement of surface sites: a Lewis acid site to stabilize the OH species adjacent to the "hopping" proton, and a Brønsted base site adjacent to the acid site to stabilize the donated proton released upon dissociation. For long-range transport by this mechanism, each Brønsted base site must also have at least one additional Lewis acid site located directly adjacent to it in a similarly favorable configuration. This Lewis acid site must in turn have an additional Brønsted base site next to it, forming a "chain" along which the proton can be passed.

Our results demonstrate that even in the absence of Lewis acid sites, water-mediated proton hopping still occurs. The process is more rapid than that observed on TiO_2(110), where the Lewis acid sites trap the water molecules, slowing the process considerably so that H diffusion events could be directly observed (in STM movies on a time scale of a few seconds) at temperatures as high as ~187 K (*15*). On FeO(111), however, the H diffusion process is too fast to be directly resolvable on a similar time scale, even at 105 K. The observed rapid proton hopping on FeO(111) in the form of the large "collective" diffusion events is related not only to the ease with which protons are exchanged but also to the combination of high stability and high mobility of water molecules adsorbed through hydrogen bonding.

Our use of both STM and DFT methods revealed that the mobility of hydrogen on the FeO(111) thin film grown on Pt(111) is completely dominated by a water-mediated hopping mechanism. This mechanism, revealed for H atoms on the fcc domains, consists of the transfer of a surface proton to a water molecule, leading to a short-lived hydronium ion species at the transition state. No bonding occurs between the water molecule and the metal cations of the FeO thin film. This domain-specific adsorption of H and the facile diffusion of H_2O between these domains on the flat FeO(111)/Pt(111) model oxide surface illustrate how a templating species (e.g., H) may be used to specifically guide the surface transport of another molecule (e.g., H_2O) with nanometer-scale precision. These nanoscale conveyor belts might be useful in emerging nanofluidics applications (*28*, *29*), including nanotube sensors and the modeling of transport across biological membranes.

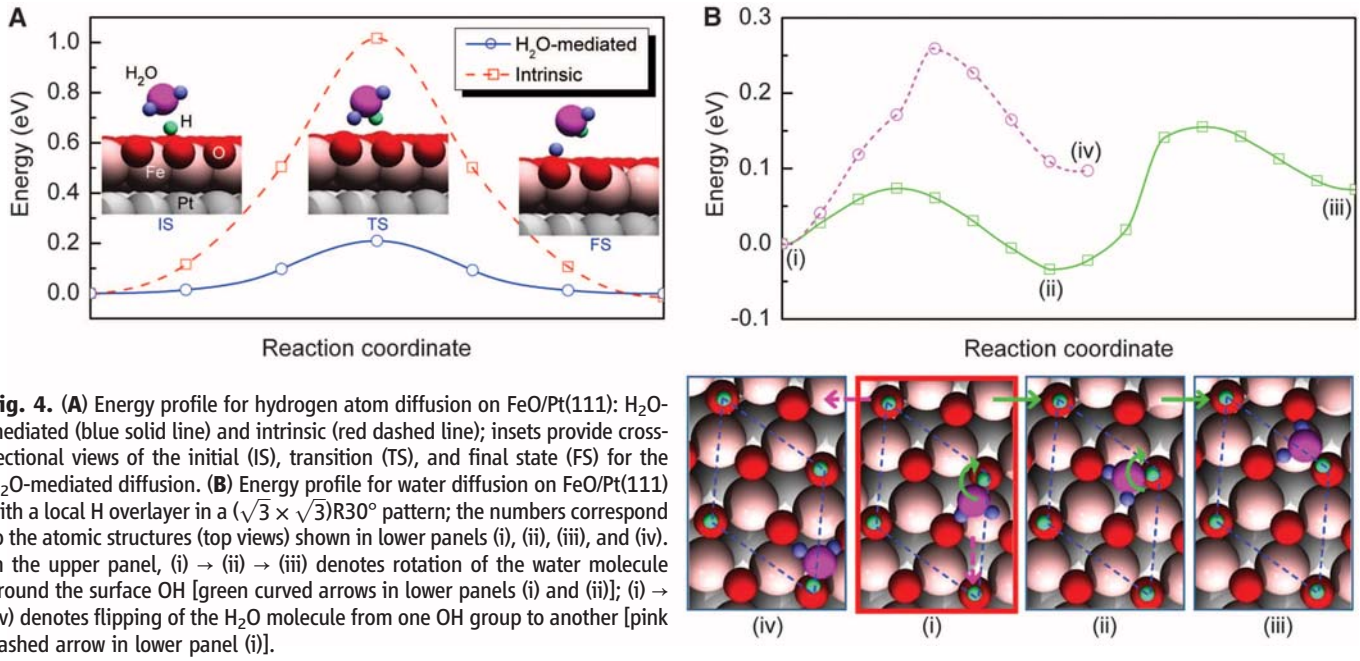


Fig. 4. (**A**) Energy profile for hydrogen atom diffusion on FeO/Pt(111): H_2O-mediated (blue solid line) and intrinsic (red dashed line); insets provide cross-sectional views of the initial (IS), transition (TS), and final state (FS) for the H_2O-mediated diffusion. (**B**) Energy profile for water diffusion on FeO/Pt(111) with a local H overlayer in a $(\sqrt{3} \times \sqrt{3})$R30° pattern; the numbers correspond to the atomic structures (top views) shown in lower panels (i), (ii), (iii), and (iv). In the upper panel, (i) → (ii) → (iii) denotes rotation of the water molecule around the surface OH [green curved arrows in lower panels (i) and (ii)]; (i) → (iv) denotes flipping of the H_2O molecule from one OH group to another [pink dashed arrow in lower panel (i)].

References and Notes

1. W. C. Conner, J. L. Falconer, *Chem. Rev.* **95**, 759 (1995).
2. V. V. Rozanov, O. V. Krylov, *Russ. Chem. Rev.* **66**, 107 (1997).
3. R. D. Cortright, R. R. Davda, J. A. Dumesic, *Nature* **418**, 964 (2002).
4. I. Ait-Ichou, M. Formenti, B. Pommier, S. J. Teichner, *J. Catal.* **91**, 293 (1985).
5. A. M. Hilmen, D. Schanke, A. Holmen, *Catal. Lett.* **38**, 143 (1996).
6. H. Cheng, L. Chen, A. C. Cooper, X. Sha, G. P. Pez, *Energy Environ. Sci.* **1**, 338 (2008).
7. K.-D. Jung, A. T. Bell, *J. Catal.* **193**, 207 (2000).
8. T. Ioannides, X. E. Verykios, *J. Catal.* **143**, 175 (1993).
9. R. B. Levy, M. Boudart, *J. Catal.* **32**, 304 (1974).
10. A. Higazy, M. Kassem, M. Sayed, *J. Phys. Chem. Solids* **53**, 549 (1992).
11. J. Kanellopoulos *et al.*, *J. Catal.* **255**, 68 (2008).
12. G. Ertl, *Angew. Chem. Int. Ed.* **47**, 3524 (2008).
13. S.-C. Li, L.-N. Chu, X.-Q. Gong, U. Diebold, *Science* **328**, 882 (2010).
14. S.-C. Li *et al.*, *J. Am. Chem. Soc.* **130**, 9080 (2008).
15. S. Wendt *et al.*, *Phys. Rev. Lett.* **96**, 066107 (2006).
16. H. C. Galloway, J. J. Benitez, M. Salmeron, *Surf. Sci.* **298**, 127 (1993).
17. J. L. Daschbach, Z. Dohnálek, S.-R. Liu, R. S. Smith, B. D. Kay, *J. Phys. Chem. B* **109**, 10362 (2005).
18. Y. Joseph, C. Kuhrs, W. Ranke, M. Ritter, W. Weiss, *Chem. Phys. Lett.* **314**, 195 (1999).
19. U. Leist, W. Ranke, K. Al-Shamery, *Phys. Chem. Chem. Phys.* **5**, 2435 (2003).
20. M. A. Henderson, *Surf. Sci. Rep.* **46**, 1 (2002).
21. W. X. Huang, W. Ranke, *Surf. Sci.* **600**, 793 (2006).
22. J. Knudsen *et al.*, *Surf. Sci.* **604**, 11 (2010).
23. See supplementary materials on *Science* Online.
24. G. Kresse, J. Furthmüller, *Phys. Rev. B* **54**, 11169 (1996).
25. J. P. Perdew, Y. Wang, *Phys. Rev. B* **45**, 13244 (1992).
26. P. E. Blöchl, *Phys. Rev. B* **50**, 17953 (1994).
27. G. Kresse, D. Joubert, *Phys. Rev. B* **59**, 1758 (1999).
28. H. Daiguji, *Nat. Nanotechnol.* **5**, 831 (2010).
29. G. M. Whitesides, *Nature* **442**, 368 (2006).

Acknowledgments: Supported by the Danish Research Agency, the Strategic Research Council, the Villum Kahn Rasmussen Foundation, the Carlsberg Foundation, the Lundbeck Foundation, Haldor Topsøe A/S, and the European Research Council through an advanced ERC grant. Work at the University of Wisconsin was supported by the U.S. Department of Energy, Basic Energy Sciences, Chemical Sciences Office. C.A.F. thanks NSF for a Graduate Research Fellowship. G.P., C.A.F., L.C.G., and M.M. thank the National Energy Research Scientific Computing Center, Pacific Northwest National Laboratory, Argonne National Laboratory, and Oak Ridge National Laboratory for computational resources and thank J. A. Dumesic for fruitful discussions.

Supplementary Materials

www.sciencemag.org/cgi/content/full/336/6083/889/DC1
Materials and Methods
Figs. S1 to S3
References (*30*–*40*)
Movies S1 to S4

23 January 2012; accepted 3 April 2012
10.1126/science.1219468

The Active Site of Methanol Synthesis over $Cu/ZnO/Al_2O_3$ Industrial Catalysts

Malte Behrens,[1]* Felix Studt,[2]* Igor Kasatkin,[1] Stefanie Kühl,[1] Michael Hävecker,[3] Frank Abild-Pedersen,[2] Stefan Zander,[1] Frank Girgsdies,[1] Patrick Kurr,[4] Benjamin-Louis Kniep,[4] Michael Tovar,[5] Richard W. Fischer,[4] Jens K. Nørskov,[2,6] Robert Schlögl[1]

One of the main stumbling blocks in developing rational design strategies for heterogeneous catalysis is that the complexity of the catalysts impairs efforts to characterize their active sites. We show how to identify the crucial atomic structure motif for the industrial $Cu/ZnO/Al_2O_3$ methanol synthesis catalyst by using a combination of experimental evidence from bulk, surface-sensitive, and imaging methods collected on real high-performance catalytic systems in combination with density functional theory calculations. The active site consists of Cu steps decorated with Zn atoms, all stabilized by a series of well-defined bulk defects and surface species that need to be present jointly for the system to work.

Methanol is produced industrially from synthesis gas mixtures ($H_2/CO_2/CO$) at elevated pressures P (50 to 100 bar) and temperatures T (200° to 300°C) over $Cu/ZnO/Al_2O_3$ catalysts, with a worldwide demand of ~50 Mtons $year^{-1}$. This catalytic system is also of interest for the potential use of methanol as a sustainable synthetic fuel obtained by hydrogenation of captured CO_2 (*1*). The phenomenological optimization of the preparation of catalytically very active "methanol copper" is far more advanced than the fundamental understanding of its high catalytic activity. The reaction mechanism of industrial methanol synthesis as well as the nature of the active site on Cu/ZnO-based high-performance catalysts have been debated (*2*) and are still not comprehensively understood. Here, we present experimental evidence for a structural model of the active site and use quantum chemical calculations to rationalize the experimentally observed structure-performance relation.

Industrial Cu/ZnO-based catalysts are prepared by a coprecipitation method (*3*) that creates porous aggregates of Cu and ZnO nanoparticles (NPs) (*4*) when Cu-rich molar compositions of Cu:Zn near 70:30 are used (*5*). The industrial system is a bulk catalyst characterized by a high Cu:Zn ratio with >50 mol% Cu (metal base), approximately spherical Cu NPs of a size around 10 nm, and ZnO NPs that are arranged in an alternating fashion to form porous aggregates. These aggregates expose a large Cu surface area of up to ~40 $m^2\ g^{-1}$. Furthermore, industrial catalysts contain low amounts of a refractory oxide as structural promoter (*6*), in most cases up to ~10% Al_2O_3. Omitting any of the constituting elements drastically reduces the performance of the system.

One important key to high performance is a large accessible Cu surface area (*7*), which has been observed to scale linearly with the activity for sample families with a similar preparation history (*8*). However, between these families considerably different intrinsic activities, that is, activities normalized by the Cu surface area, can be found. Thus, different "qualities" of Cu surfaces can be prepared that vary in the activity of their active sites and/or in the concentration of these sites. Hence, methanol synthesis over Cu appears to be a structure-sensitive reaction. Single-crystal studies (*9*–*11*) report turnover frequencies (TOF) for methanol synthesis ranging from as low as 1.3×10^{-6} to $6 \times 10^{-3}\ s^{-1}$ per site.

ZnO functions as a physical spacer between Cu NPs and helps disperse the Cu phase in the course of catalyst preparation (*5*) and is thus responsible for the high Cu surface areas of industrial catalysts. However, the presence of ZnO increases the intrinsic activity of Cu-based methanol synthesis catalysts, an effect known as the Cu-ZnO synergy (*3*, *12*, *13*), and has led to many different (and conflicting) mechanistic models (*14*–*20*). This situation is partially the result of some models mainly having their bases in results from simplified samples, ranging from Cu single crystals to Cu NPs supported on highly crystalline ZnO with a low loading, that have compositions and microstructure that strongly deviate from that of the industrial catalyst described above.

Investigations on industrial samples and other coprecipitated systems have suggested that defects (*4*) and lattice strain (*21*) in the Cu NPs affect the intrinsic activity of the Cu surface. To study the role of defects in the real $Cu/ZnO/(Al_2O_3)$ composite system, we prepared a series of five functional catalysts and compared them to a pure Cu metal reference sample; details on the different samples can be found as

[1]Fritz-Haber-Institut der Max-Planck-Gesellschaft, Department of Inorganic Chemistry, Faradayweg 4-6, 14195 Berlin, Germany. [2]SUNCAT Center for Interface Science and Catalysis, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA. [3]Division Solar Energy Research, Elektronenspeicherring BESSY II, Helmholtz-Zentrum Berlin für Materialien und Energie, Albert-Einstein-Strasse 15, 12489 Berlin, Germany. [4]Süd-Chemie AG, Research and Development Catalysts, Waldheimer Straße 13, 83052 Bruckmühl, Germany. [5]Institute for Complex Magnetic Materials, Helmholtz-Zentrum Berlin für Materialien und Energie, Hahn-Meitner-Platz 1, 14109 Berlin, Germany. [6]Department of Chemical Engineering, Stanford University, Stanford, CA 94305, USA.

*To whom correspondence should be addressed. E-mail: behrens@fhi-berlin.mpg.de (M.B.); studt@slac.stanford.edu (F.S.)

supplementary information (tables S1 and S2) (*22*). All samples had a high Cu loading, Cu NPs sizes between 5 and 15 nm, and exposed Cu surface areas of 10 $m^2\ g^{-1}$ or greater. These properties made them similar to industrial catalysts. In order to allow for a reliable correlation of catalytic and structural data, the microstructural homogeneity of the prepared catalysts was carefully checked. Samples were prepared from nearly single-phase precursor materials, which resulted in relatively homogeneous element distributions and monomodal Cu NP size distributions after thermal treatment (figs. S1 to S3) (*22*). The catalytic activity of all six samples (Fig. 1A) was measured under industrial conditions at $P = 60$ bar and $T = 210°$ and 250°C in a typical syngas mixture (*22*). The ZnO-free Cu reference exhibited little activity, whereas the catalysts that performed best were prepared following the industrial synthesis method. Dividing the performance by the Cu surface areas results in the intrinsic activities, which are shown in Fig. 1B normalized to the intrinsically most active catalyst for each temperature. The scatter of the data shows that the Cu surface area alone cannot explain the differences in performance. The use of the full exposed Cu surface area for a formal calculation of TOFs resulted in values varying from 5.4×10^{-4} for the pure Cu reference to $2.1 \times 10^{-2}\ s^{-1}$ per site for the intrinsically most active samples; this value is at the medium to higher end of the reported values (*9*–*11*).

To find a structural explanation for the observed trend, we performed neutron diffraction experiments on the reduced catalysts. Broad peaks of the metallic Cu face-centered cubic (fcc) phase indicative of small crystallite domains (3.8 to 9.9 nm) were present in all catalyst samples (tables S3 and S4 and fig. S4) (*22*). The inactive pure Cu reference sample exhibited sharper peaks and larger domains of >100 nm. We performed an analysis of the planar defect structure by using a pattern decomposition method (table S5) (*22*). Characteristic diffraction peaks will broaden and shift from their ideal position as a function of increased stacking fault concentration (*23*). The shift of the 111 and 200 peaks toward each other and the simultaneous shift of the higher-order peaks 222 and 400 away from each other are especially characteristic of stacking faults. The neutron scattering data allowed for a sufficiently reliable fitting of the 400 peak position of the nanostructured Cu phase. The ratios $d_{hhh}/d_{(2h)00}$ for $h = 1, 2$, which are expected to be constant at $2/\sqrt{3} = 1.1547$ for an ideal fcc structure, are shown in Fig. 1C. For the inactive pure Cu sample, both ratios fall near the expected ideal value, whereas the catalytically active materials showed a lower value for $h = 1$ and a higher one for $h = 2$, which is consistent with the presence of stacking faults in the Cu particles. For quantification (Fig. 1C), we used the spacing of the more intensive 111 and 200 peaks (*24*); the resulting stacking fault density (Fig. 1D) scales linearly with the intrinsic activity. The highly active methanol copper is a defective form of nanoparticulate Cu rich in planar defects like stacking faults. The high abundance of defects in the active materials is tentatively related to the confined crystallization of the Cu NPs in strong interfacial contact with the ZnO component during the mild catalyst activation procedure, leading to a kinetically trapped form of Cu.

The nonideal nature of active Cu is also manifested as microstrain, to which lattice defects will contribute. The absolute amount of strain, although not negligible, is not large enough to cause substantial changes in the binding energies or barriers. Typically, defects appear as a mechanism of strain relaxation, and some residual strain is concentrated around them. Thus, defects can be considered as being coupled to strain (*22*). Accordingly, we found in our series of samples a coarse trend of the intrinsic activity with higher lattice strain (fig. S5) (*22*). The general importance of strain for Cu/ZnO catalysts has been highlighted before (*4*, *21*). We suggest that the main role of the bulk defects for catalysis is that an extended defect induces a line defect at the exposed surfaces—typically a step, as can be observed in Fig. 1D, which shows how a stacking

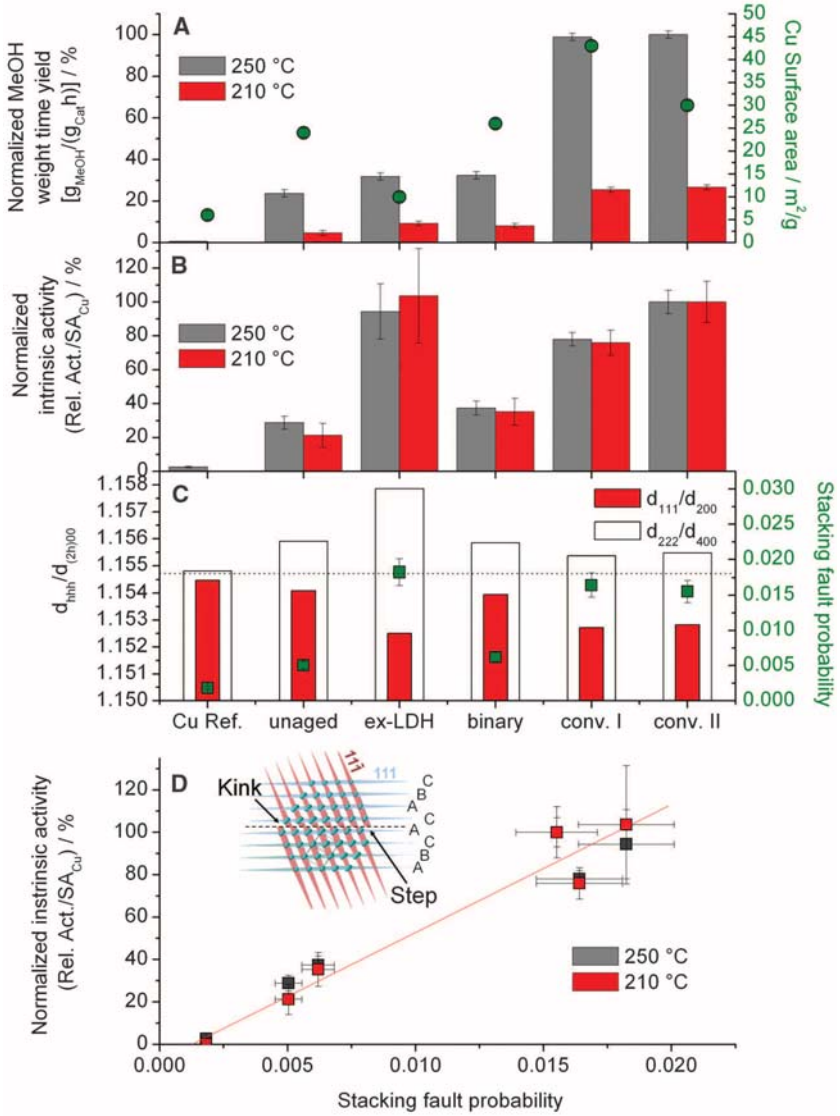


Fig. 1. (**A**) Catalytic activities and Cu surface areas of the Cu reference material and the five $Cu/ZnO/Al_2O_3$ catalysts in methanol synthesis ($P = 60$ bar, $T = 210°$, 250°C, normalized to the most active sample). (**B**) Intrinsic activities per Cu surface area obtained after dividing by the Cu surface area (normalized: most active sample = 100% at each temperature). (**C**) Deviation of d_{111}/d_{200} and d_{222}/d_{400} observed in the neutron diffraction patterns and resulting stacking fault probabilities of the Cu particles. The dashed line refers to the ideal fcc structure. For a detailed description of the samples and explanation of sample labeling, see (*22*). (**D**) Relation of the intrinsic activity of Cu to the concentration of stacking faults. (Inset) Schematic of how a stacking fault in 111 can generate kinks and surface steps in the $1\bar{1}\bar{1}$ facet. Error bars indicate uncertainties determined on basis of replicate measurements (catalytic activity and copper surface area). The error bar of the diffraction peak analysis is based on an estimated uncertainty of 0.1% for the angular peak position.

fault in 111 creates a step on the $11\bar{1}$ surface of a Cu crystallite. A twin boundary terminating at a surface is associated with a kink.

The effect of steps at the Cu surface on the catalytic properties was also confirmed by density functional theory (DFT) calculations on different Cu surfaces. In this study, DFT was used to provide qualitative confirmation of the experimentally observed trends and to rationalize the effect of the structural features that have been identified to be relevant for the catalytic properties of the catalyst's surface. To attain some independence of presumptions on the reaction mechanism, we studied methanol formation from both CO_2 and CO (*25*). A flat Cu(111) surface represents the ideal defect-free catalyst, whereas a stepped Cu(211) surface was used to include the effect of surface defects (Fig. 2, black and blue curves). Figure 2B shows the CO_2 hydrogenation pathway on the two different surfaces. For clarity, only the lowest-energy pathway is shown, which is the same for both surfaces. Energetics of the other intermediates are given in tables S6 to S8 (*22*). The barrier for the splitting of molecular hydrogen was calculated to be 0.74 and 0.84 eV on the Cu(211) and Cu(111) surfaces, respectively, so surface hydrogen was readily available under reaction conditions. Hydrogenation of CO_2 proceeded via formation of HCOO, HCOOH, and H_2COOH. The C-O bond of H_2COOH was split to yield adsorbed H_2CO and OH, where H_2CO is hydrogenated to methanol via the methoxy (CH_3O) intermediate. Surface OH was removed as water. A similar pathway for CO_2 hydrogenation on the (111) surface of Cu has been suggested recently (*26*). Other theoretical studies of this reaction have considered Cu(100) (*27*), small Cu clusters (*28*), or Zn atoms deposited on Cu(111) (*29*).

As shown in Fig. 2B, the flat Cu(111) surface bound the intermediates more weakly than did Cu(211). Essentially all intermediates are thermodynamically less stable than CO_2 and H_2 in the gas phase. Note that these thermodynamics include effects of pressure; high CO_2 and H_2 pressure strengthen adsorption energies and make the formation of methanol and water downhill in energy, explaining why high pressures are needed for this process. Both the energies of the intermediates and the transition-state were stabilized considerably for the (211) surface compared with the (111) surface, rendering the steps more active than the terraces. Similar results were obtained for hydrogenation of CO to methanol (Fig. 2C). CO hydrogenation proceeded via hydrogenation of the carbon atom of CO, with the intermediates being HCO, H_2CO, and H_3CO. H_3CO was then hydrogenated to methanol. The last two intermediates are the same as for the hydrogenation of CO_2. Steps again lower the adsorption energies of the intermediates substantially compared with the flat surface.

The stepped Cu(211) surface or the stacking fault-created step shown in the inset of Fig. 1D can be regarded as model situations. We also examined the relation of bulk defects and surface steps in the most active catalyst with aberration-corrected high-resolution transition electron microscopy (HRTEM). The vast majority of the investigated Cu NPs were faulted and exhibited planar extended defects, stacking faults, and twin boundaries that ran through the whole particle. It can be seen from the micrographs in Fig. 3 that the NP shapes can be generally approximated by a sphere. The curvature of the particle causes the surface to intrinsically contain a number of steps. The HRTEM image in Fig. 3A shows stepped surface facets like (211) and (522) being responsible for the curvature at the lower exposed side of the Cu NP. Surface faceting also changed along the line where the twin boundary terminates at the surface. This kind of kink is associated with an inward curvature of the surface, which does not occur on regular spherical or ellipsoidal fcc particles or Wulff polyhedra. Another example (Fig. 3B) showed a pattern of planar bulk defects (twin boundaries), as reflected in changes of the surface faceting creating a local inward curvature of the NP. Despite the absence of stepped surface facets at this part of the NP, a number of steps were created by the inward kinks, where the twin boundaries meet the surface. Figure 3C shows that twin boundaries could create distinctive surface ensembles even if the Cu surface of a larger NP appears essentially flat. The change of the surface faceting

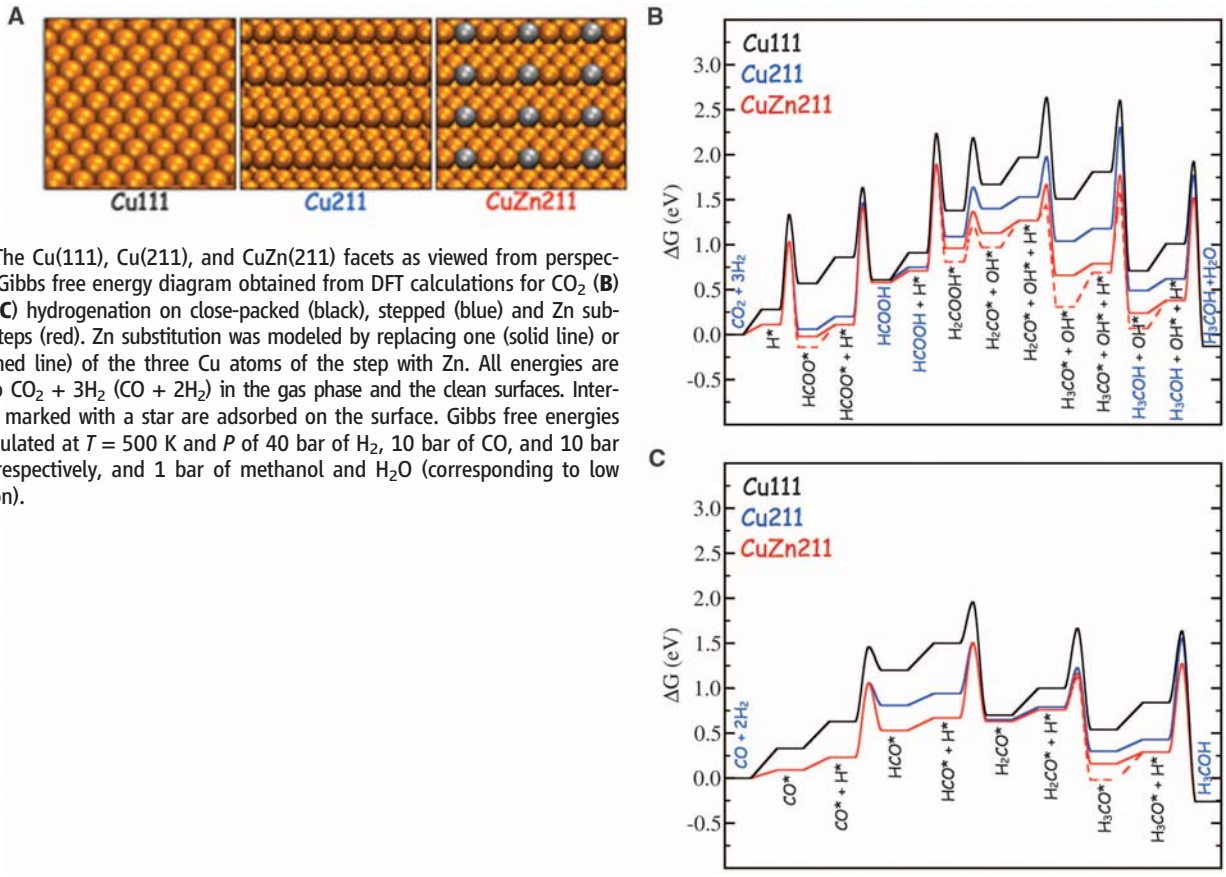


Fig. 2. The Cu(111), Cu(211), and CuZn(211) facets as viewed from perspective (**A**). Gibbs free energy diagram obtained from DFT calculations for CO_2 (**B**) and CO (**C**) hydrogenation on close-packed (black), stepped (blue) and Zn substituted steps (red). Zn substitution was modeled by replacing one (solid line) or two (dashed line) of the three Cu atoms of the step with Zn. All energies are relative to $CO_2 + 3H_2$ ($CO + 2H_2$) in the gas phase and the clean surfaces. Intermediates marked with a star are adsorbed on the surface. Gibbs free energies were calculated at $T = 500$ K and P of 40 bar of H_2, 10 bar of CO, and 10 bar of CO_2, respectively, and 1 bar of methanol and H_2O (corresponding to low conversion).

from (111) to (100) is associated only with a slightly obtuse angle near 180°. Neighboring to the position of the kink, a column of surface atoms was observed whose position stuck out of the regular surface (Fig. 3D, arrow). Again, such an arrangement can be described as a high-energy site created by the termination of a planar defect at the surface of the Cu NP.

Because the presence of a defect at the active site alone does not directly involve ZnO, it cannot explain the observation (*12*, *13*) of Cu-ZnO synergy in physical mixtures. It seems likely that the Cu-ZnO synergy is related to strong metal support interaction (SMSI) between Cu and ZnO leading to a partial coverage of the Cu surface with ZnO_x under reducing conditions. SMSI has been observed on Cu/ZnO-based catalysts by using vibrational spectroscopy (*30*) and thermal desorption of probe molecules (*31*) and by monitoring the wetting behavior of Cu/ZnO model catalysts (*32*). In the high-performance catalyst studied here, the presence of a disordered overlayer of the Cu NPs with a thickness of about 1 nm can be seen in some HRTEM images (Fig. 3, B and C). In the complex real catalyst, fully covered NPs coexist with partially covered and practically uncovered ones. To identify the layer as ZnO_x, we investigated the surface composition of the most active catalyst by using ambient pressure x-ray photoemission spectroscopy (XPS). In agreement with previously reported data (*16*, *33*), the Cu:Zn ratio at the catalyst's surface dropped during activation in hydrogen. For our catalyst, the Cu:Zn ratio was inverted from its nominal value of 70:30 (calcined) to ~30:70 (reduced), supporting the idea of an SMSI effect (Fig. 4A). The (partial) ZnO_x coverage of the surface of the reduced Cu NPs in the catalyst was also evidenced by tuning the information depth of the experiment from about 0.6 to 2.3 nm by variation of the kinetic energy of the incoming x-ray beam (Fig. 4B). Core level fitting of the Zn 3p and Cu 3p signals showed the enrichment of Zn at the surface of the catalyst (fig. S6) (*22*), whereas with a higher information depth the Cu:Zn ratio is slowly approaching toward the nominal composition. The calcined catalyst did not show any surface enrichment of Zn, and the effect was fully reversible upon recalcination of the catalyst. Given the dynamics of the ZnO component in this Cu/ZnO catalyst and the observation of surface decoration of the Cu NPs already at relatively mild conditions of low partial pressure of hydrogen, further progression of this effect under strongly reducing conditions may lead to formation of a CuZn surface alloy, as discussed by several authors (*3*, *30*, *32*).

The beneficial role of Zn at the catalyst's surface can be explained by DFT calculations. To incorporate the effect of Zn, we studied a Cu(211) surface where Cu in the step is partially substituted by Zn (Fig. 2, red curves). Alloying of Zn into the Cu step further increased the adsorption strength of HCO, H_2CO, and H_3CO and decreased the barriers. Hence, the rate of methanol synthesis was further increased. The order of activity for CO_2 as well as for CO hydrogenation is CuZn(211) > Cu(211) > Cu(111). The most active surface was therefore found to be a Cu step with Zn alloyed into it. The adsorption properties of alloyed CuZn(111) surface have been experimentally observed to be modified from those of pure Cu(111) (*18*). Indeed, species that are bound to the surface through oxygen atoms, such as formate and hydroxyl,

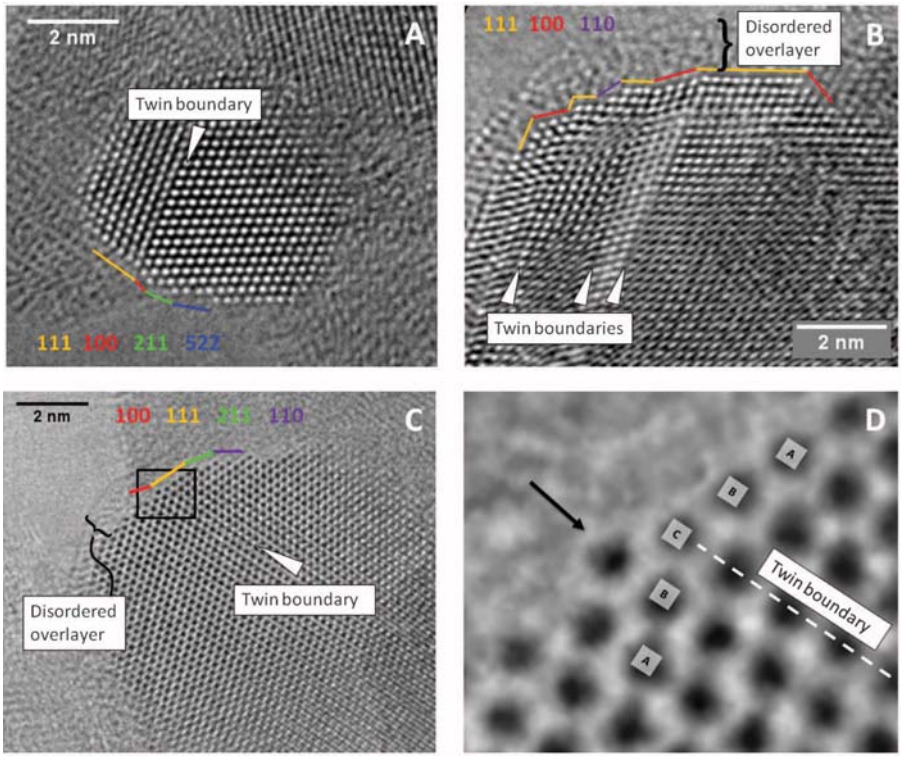


Fig. 3. (**A** to **D**) Aberration-corrected HRTEM images of Cu particles in the conventionally prepared, most-active $Cu/ZnO/Al_2O_3$ catalyst. (D) is a close-up of the marked area in (C).

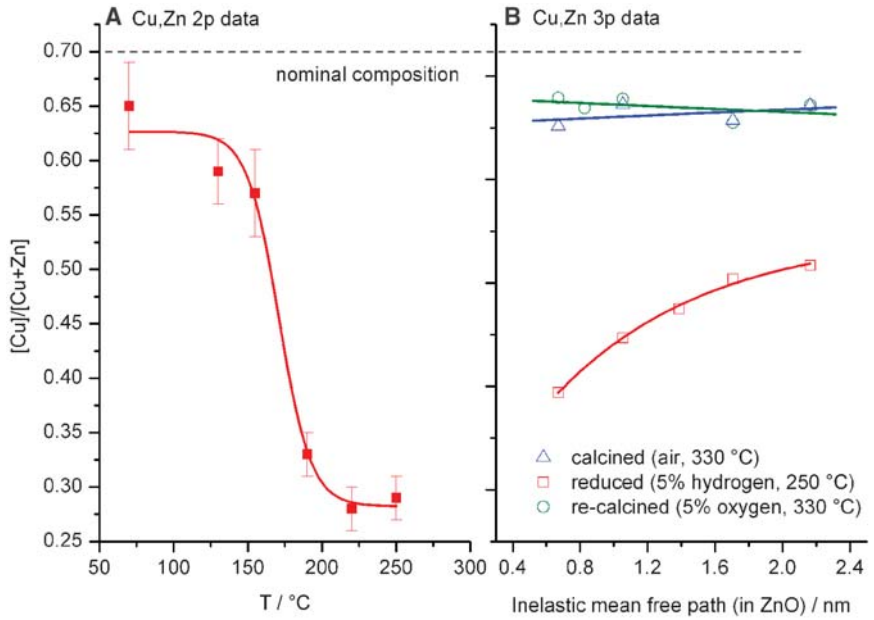


Fig. 4. Surface and near-surface composition of the most active $Cu/ZnO/Al_2O_3$ catalyst of this study with a Cu:Zn ratio of 70:30 recorded with synchrotron XPS. (**A**) In situ Cu, Zn 2p data during reduction in 0.25 mbar hydrogen with a heating rate of 2 K min^{-1}. Error bars represent estimated uncertainty based on several fits with random variation of background fitting parameters. (**B**) Environmental Cu, Zn 3p data of the calcined, the pre-reduced (in 5% hydrogen at 250°C) and recalcined (5% oxygen at 330°C) catalyst as a function of information depth. Lines are guides to the eye.

will widely cover the surface of the catalyst under methanol synthesis conditions. If more Zn atoms are considered in the CuZn(211) surface, the binding to these species is further strengthened (Fig. 2). According to this trend and in agreement with the higher oxophilicity of Zn compared with Cu, these species will bind to the surface via Zn atoms, leading to a formal oxidation of the Zn component. Thus, under steady-state conditions, the oxidation state of Zn is adjusted to a partially oxidized $Zn^{\delta+}$ state, which can be formed by reduction from the ZnO particles through SMSI as well as from a CuZn surface alloy by adsorbate-induced oxidation. The unique role of ZnO in the industrial catalyst is probably related to the stability of this intermediate oxidation state under the reducing potential of methanol synthesis conditions. Its reducibility is high enough to allow for partial reduction but sufficiently low not to favor bulk alloying. Other promoters that bind oxygen in the same range as Zn may have a similar effect. The Cu/ZrO_2 system, for example, is also an active methanol synthesis catalyst (*34*).

Combining the experimental and theoretical results, a model for the active site of methanol synthesis over industrial catalysts emerges. Undistorted pure Cu was quite inactive in the methanol synthesis experiment. The same result was obtained for the flat Cu(111) surface in the DFT calculations. High activity was generated by two factors. First, the presence of steps at the Cu surface is required, which can be stabilized by bulk defects like stacking faults or twin boundaries terminating at the surface. The increase in activity is explained by a stronger binding of the intermediates on stepped sites and lower energy barriers between them. The bulk defect structure in the real catalyst is a result of a well-optimized low-temperature preparation method.

The second requirement is the presence of $Zn^{\delta+}$ at the defective (stepped) Cu surface, which in the high-performance catalyst is a result of a dynamic SMSI effect leading to partial coverage of the metal particles with ZnO_x. Substitution of Zn into the Cu steps further strengthens the binding of the intermediates and increases the activity of the catalyst. The data presented suggest that the presence of steps and their close proximity to ZnO_x on the surface of the Cu particles create the ensemble needed to render the very active methanol copper: a Cu step with a nearby Zn serving as adsorption site for oxygen-bound intermediates.

These two requirements are fulfilled only for a small and varying fraction of the metallic Cu surface area, explaining the differences in intrinsic activity observed here and also in literature. Thus, under industrially relevant conditions a small fraction of the surface is largely contributing to the activity, which cannot be easily mimicked by simplified model approaches. We propose that the TOF of this reaction channel should be considerably higher compared with the values calculated on the basis of the full exposed Cu surface area.

References and Notes

1. G. A. Olah, A. Goeppert, G. K. Surya Prakash, *Beyond Oil and Gas: The Methanol Economy* (Wiley-VCH, Weinheim, Germany, 2006).
2. J. B. Hansen, P. E. Højlund Nielsen, in *Handbook of Heterogenous Catalysis*, G. Ertl, H. Knözinger, F. Schüth, J. Weitkamp, Eds. (Wiley-VCH, Weinheim, Germany, ed. 2, 2008), pp. 2920–2949.
3. M. S. Spencer, *Top. Catal.* **8**, 259 (1999).
4. I. Kasatkin, P. Kurr, B. Kniep, A. Trunschke, R. Schlögl, *Angew. Chem.* **119**, 7465 (2007).
5. M. Behrens, *J. Catal.* **267**, 24 (2009).
6. M. Kurtz, H. Wilmer, T. Genger, O. Hinrichsen, M. Muhler, *Catal. Lett.* **86**, 77 (2003).
7. The exposed Cu surface area can be measured using reactive chemisorption of N_2O, which leads to a surface oxidation of Cu ideally yielding a Cu_2O monolayer.
8. M. Kurtz *et al.*, *Catal. Lett.* **92**, 49 (2004).
9. J. Yoshihara, C. T. Campbell, *J. Catal.* **161**, 776 (1996).
10. P. B. Rasmussen *et al.*, *Catal. Lett.* **26**, 373 (1994).
11. J. Szanyi, D. W. Goodman, *Catal. Lett.* **10**, 383 (1991).
12. R. Burch, S. E. Golunski, M. S. Spencer, *J. Chem. Soc. Faraday Trans.* **86**, 2683 (1990).
13. Y. Kanai, T. Watanabe, T. Fujitani, T. Uchijima, J. Nakamura, *Catal. Today* **28**, 223 (1996).
14. K. Klier, *Adv. Catal.* **31**, 243 (1982).
15. V. Ponec, *Surf. Sci.* **272**, 111 (1992).
16. W. P. A. Jansen *et al.*, *J. Catal.* **210**, 229 (2002).
17. J. C. Frost, *Nature* **334**, 577 (1988).
18. J. Nakamura, Y. Choi, T. Fujitani, *Top. Catal.* **22**, 277 (2003).
19. K. C. Waugh, *Catal. Today* **15**, 51 (1992).
20. P. L. Hansen *et al.*, *Science* **295**, 2053 (2002).
21. M. M. Günter *et al.*, *Catal. Lett.* **71**, 37 (2001).
22. See supplementary materials on *Science* Online.
23. This effect arises from the generation of thin hexagonal domains in the cubic lattice with the change in stacking sequence of the hexagonally close-packed (111) layers at the stacking fault (ideal is A-B-C-A; stacking fault, A-B-C-B-C-A; twin boundaries, A-B-C-B-A). For more details and quantitative treatment see, e.g., (*24*).
24. B. E. Warren, *X-ray Diffraction* (Dover, New York, 1990).
25. G. C. Chinchen, P. J. Denny, J. R. Jennings, M. S. Spencer, K. C. Waugh, *Appl. Catal.* **36**, 1 (1988).
26. L. C. Grabow, M. Mavrikakis, *ACS Catal.* **1**, 365 (2011).
27. Z.-M. Hu, K. Takahashi, H. Nakatsuji, *Surf. Sci.* **442**, 90 (1999).
28. Y. Yang, J. Evans, J. A. Rodriguez, M. G. White, P. Liu, *Phys. Chem. Chem. Phys.* **12**, 9909 (2010).
29. Y. Morikawa, K. Iwata, K. Terakura, *Appl. Surf. Sci.* **169–170**, 11 (2001).
30. N.-Y. Topsøe, H. Topsøe, *Top. Catal.* **8**, 267 (1999).
31. R. Naumann d'Alnoncourt *et al.*, *Phys. Chem. Chem. Phys.* **8**, 1525 (2006).
32. J. D. Grunwaldt, A. M. Molenbroek, N. Y. Topsoe, H. Topsoe, B. S. Clausen, *J. Catal.* **194**, 452 (2000).
33. J. Słoczynski *et al.*, *Appl. Catal. A Gen.* **310**, 127 (2006).
34. I. A. Fisher, H. C. Woo, A. T. Bell, *Catal. Lett.* **44**, 11 (1997).

Acknowledgments: We thank M. Muhler and O. Hinrichsen for fruitful discussions. M.B., I.K., S.K., S.Z. F.G., and R.S. acknowledge the Bundesministerium für Bildung und Forschung (FKZ 01RI0529) and Süd-Chemie AG for financial support. P.K., B.-L.K., and R.W.F. thank the Bayerisches Wissenschaftsministerium (NW-0810-0002) for financial support of this work. F.S., F.A.-P., and J.K.N. wish to acknowledge support from the (U.S.) Department of Energy, Office of Basic Energy Sciences. The data presented in this work can be obtained from the corresponding authors upon request.

Supplementary Materials

www.sciencemag.org/cgi/content/full/science.1219831/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S6
Tables S1 to S8
References (*35–52*)

30 January 2012; accepted 3 April 2012
Published online 19 April 2012;
10.1126/science.1219831

Structures of Cage, Prism, and Book Isomers of Water Hexamer from Broadband Rotational Spectroscopy

Cristóbal Pérez,[1] Matt T. Muckle,[1] Daniel P. Zaleski,[1] Nathan A. Seifert,[1] Berhane Temelso,[2] George C. Shields,[2]* Zbigniew Kisiel,[3]* Brooks H. Pate[1]*

Theory predicts the water hexamer to be the smallest water cluster with a three-dimensional hydrogen-bonding network as its minimum energy structure. There are several possible low-energy isomers, and calculations with different methods and basis sets assign them different relative stabilities. Previous experimental work has provided evidence for the cage, book, and cyclic isomers, but no experiment has identified multiple coexisting structures. Here, we report that broadband rotational spectroscopy in a pulsed supersonic expansion unambiguously identifies all three isomers; we determined their oxygen framework structures by means of oxygen-18–substituted water ($H_2{}^{18}O$). Relative isomer populations at different expansion conditions establish that the cage isomer is the minimum energy structure. Rotational spectra consistent with predicted heptamer and nonamer structures have also been identified.

The intermolecular hydrogen-bonding interactions of water are responsible for many remarkable physical properties of the liquid and solid phases of the compound and furthermore play a pivotal role in solution chemistry and biochemistry. As a result, the accurate description of the water intermolecular potential is one of the most important problems in chemistry (*1*). One key method for quantitative analysis of water interactions is the size-selective study of the structures of water clusters (*2–5*). This problem has been attacked using several state-of-the-art techniques, including far-infrared (FIR) spectroscopy (*6–9*), helium nanodroplet isolation (HENDI) spectroscopy (*10*), infrared spectroscopy of size-selected molecular beams (*11*), molecular tagging ion-dip infrared spectroscopy (*12*, *13*), and argon-mediated, population-modulated attachment spectroscopy (*14*). Here, we report chirped-pulse Fourier transform microwave

(CP-FTMW) spectra (*15*) of clusters produced in a pulsed supersonic expansion. We thereby identify three isomers of the water hexamer, obtain direct structural information on these isomers through the analysis of isotopically labeled clusters, and establish that the cage structure is the global minimum.

The water hexamer has become a benchmark problem for computational chemistry (*16*–*22*). Theoretical studies of water cluster structures indicate that the hexamer is a special cluster size wherein three-dimensional structures become more stable than the planar ring structures found for the smaller clusters (*2*, *3*), a result supported by experimental studies of size-specific clusters of water on benzene (*12*). For this transitional cluster, there is a complex, multidimensional potential energy surface, and several low-energy structural isomers are predicted, including prism, cage, book, bag, and cyclic forms. There is consensus among the most recent calculations that the prism, cage, and book isomers are the three lowest energy isomers, with an energy separation of 1 kJ/mol or less when zero-point corrections are included (*16*, *20*). The prism has been calculated to be the lowest-energy isomer, although the prism and cage are nearly isoenergetic, and the book is slightly higher in energy.

Experimental studies of the water hexamer have identified the cage, book, and cyclic isomers based on the comparison of calculated and observed spectral properties. The hitherto most definitive structure identification has been the observation of the cage by high-resolution FIR spectroscopy (*9*). The experimental rotational constants, which measure the mass distribution in the cluster through the principal moments of inertia, were consistent only with theoretical calculations of the cage structure. The tunneling splitting patterns and dipole moment of the cluster were subsequently shown to be consistent with the theoretical predictions for the cage structure as well (*23*). No other isomers were detected. A higher-energy cyclic isomer of the water hexamer, produced under the unique cluster growth conditions of helium nanodroplets, was identified by HENDI spectroscopy (*10*) from the frequency of a vibrational band of the O-H stretch that followed the theoretical prediction for red shift as a function of cluster size. In two experiments in which the infrared spectrum in the region of the O-H stretch was measured for water hexamer clusters produced in a supersonic expansion (*11*, *14*), the comparison to theoretical spectra indicated the presence of the book isomer. In the case of the study using size-selected molecular beams (*11*), the dominance of the book isomer, which is entropically favored over the cage and prism isomers (*16*), was consistent with the warmer temperature of the sample (40 K) in these measurements compared with the slit-jet FIR spectroscopy measurements (6 K) and the HENDI experiments (0.38 K). However, recent theoretical calculations of hexamer vibrational spectra have suggested that the assignment of the book isomer may need to be revisited (*24*).

The structures of water clusters have presently been characterized by analysis of their rotational spectra. Molecular rotational spectroscopy, applicable to polar molecules in the gas phase, is one of the most exacting experimental techniques for characterizing molecular structure. Analysis of the rotational spectrum permits a high-accuracy determination of the principal moments of inertia. This accuracy is needed for structure determination using isotopic substitution because the small differences in the moments of inertia of the naturally abundant and isotopically labeled analogs are used to obtain the atom positions. The high spectral resolution of pulsed-jet microwave spectroscopy is crucial for structure determination by isotopic substitution, where it is necessary to fully resolve the several closely spaced rotational spectra of each isotopolog in the sample. The rotational spectra were measured in a CP-FTMW spectrometer with frequency coverage from 7 to 18 GHz (*15*). Separate broadband rotational spectra were recorded by flowing argon, neon, and helium over an external reservoir of water at room temperature. In addition, a measurement using neon as the carrier gas was performed using a water sample that contained 15% $H_2{}^{18}O$. The CP-FTMW data and detailed analysis of all spectra are provided in the supplementary materials.

The broadband rotational spectrum of water is shown in Fig. 1. The rotational spectrum of the water dimer (*25*) is the most intense contributor. The minimum energy structures of the water trimer, tetramer, and pentamer are effectively nonpolar in their ground states, so these clusters cannot be observed. The previously identified cage isomer of the water hexamer (*9*) gives the next strongest spectrum, observed at about a 200:1 signal-to-noise ratio in neon. Spectra for the prism and book isomers are about half as intense. The prism isomer is the only one to show fine structure associated with the rearrangement of the hydrogen-bond network through tunneling. We have also assigned rotational spectra that we attribute to a water heptamer and a water nonamer cluster, based on comparison to the rotational constants from theoretical structures (*16*). The structures are tentatively identified as the heptamer PR1 and nonamer D2dDD in the nomenclature of (*16*). The two most stable cubic structures of the octamer are nonpolar (*13*). A summary of the spectroscopic constants of these water clusters and comparison to the rotational constants predicted by electronic structure theory are given in Table 1. For the prism, the rotational constants reported in Table 1 are obtained from a fit of the center frequencies of the observed tunneling fine structure.

The rotational spectrum was measured using different inert carrier gases to gain information about the relative stability of the water hexamer isomers based on their population in the molecular expansion, as shown in Fig. 2. The electric dipole moment of the cluster is required to convert the spectral intensities to populations and

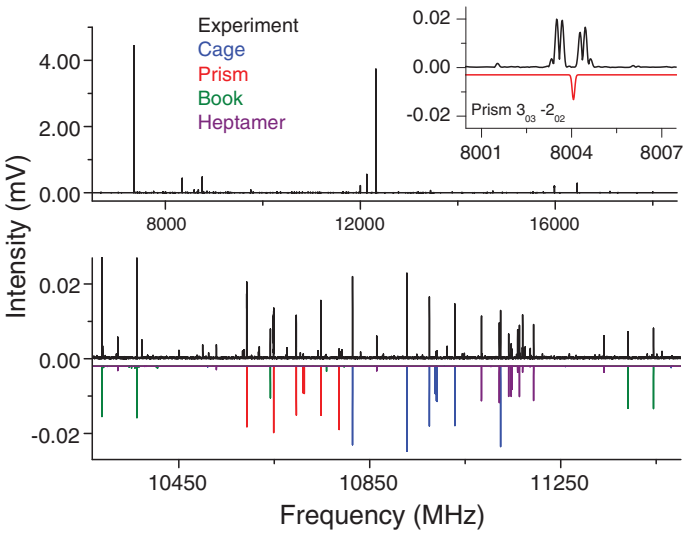


Fig. 1. The black traces show different views of the experimental CP-FTMW pulsed-jet spectrum of water using neon backing gas. The final spectrum is obtained from the signal average of 650,000 free induction decay (FID) measurements with a 20-μs gate duration. The colored traces in the bottom panel are 1.5 K (rotational temperature) simulations of the hexamer cage (blue), hexamer prism (red), hexamer book (green), and heptamer (purple). The nonamer is not seen on this scale but can be found in fig. S6. The inset in the top panel shows the 3_{03}-2_{02} rotational transition (defined using the standard nomenclature for the quantized rotational energy levels of an asymmetric top denoted J_{KaKc}, where J is the quantum number for the total rotational angular momentum and K_a and K_c are the quantum numbers for the projection of total rotational angular momentum onto the a- and c-principal rotation axes, respectively) of the prism isomer, highlighting the structure due to quantum tunneling. The simulation of the prism, shown in red, has been fit to the line centers.

[1]Department of Chemistry, University of Virginia, McCormick Road, Charlottesville, VA 22903, USA. [2]Dean's Office, College of Arts and Sciences and Department of Chemistry, Bucknell University, Lewisburg, PA 17837, USA. [3]Institute of Physics, Polish Academy of Sciences, Aleja Lotników 32/46, 02-668 Warszawa, Poland.

*To whom correspondence should be addressed. E-mail: brookspate@virginia.edu (B.H.P.); george.shields@bucknell.edu (G.C.S.); kisiel@ifpan.edu.pl (Z.K.)

the analysis uses ab initio estimates of the dipole moments reported in table S41. The estimated population ratio for cage:prism:book is 1:1:0.25 in neon and helium. However, in an argon expansion, only the cage isomer is observed (with a signal-to-noise ratio of 50:1), giving strong evidence that the cage is the global minimum energy structure for the water hexamer, as inferred in the previous FIR spectroscopy work (*9*). The mechanism for the enhanced cooling efficiency of heavier carrier gases in free-jet expansions has been recently discussed (*26*). The prism, often found to be the lowest-energy isomer in calculations using high-level quantum chemistry (*3*, *16*, *18*–*22*), is the next-lowest-energy isomer based on this population analysis. The spectra attributed to heptamer and nonamer water clusters are observed in all three carrier gases, indicating that these are the global minimum configurations.

Although the agreement between experimental and theoretical spectroscopic constants provides compelling support for the identification of the cage, prism, and book isomers, the study of isotopically labeled water clusters gives unambiguous structure identification. As shown in Fig. 2, using an $H_2{}^{18}O$-enriched sample leads to the assignment of six separate rotational spectra, each with a single $H_2{}^{18}O$ molecule incorporated into a structurally unique site of the cluster. For the prism, only six distinct spectra were identified, and none of them show the fine structure observed in the normal $H_2{}^{16}O$ cluster spectrum. This result supports the attribution of this fine structure to tunneling between several identical minima of the potential energy surface. Upon substitution by a single $H_2{}^{18}O$ molecule in the cluster, these minima are no longer energetically equivalent and the quantum tunneling is quenched. One implication of this result is that the tunneling dynamics observed for the prism are not the bifurcation pathways that are the focus of previous analysis of FIR spectra because these motions would still connect equivalent minima of the potential energy surface upon single ^{18}O substitution (*23*).

The initial structural analysis of the water hexamers uses the method of Kraitchman, in which the oxygen atom framework is built up atom by atom by converting the changes in the moments of inertia upon isotopic substitution into atom coordinates in the principal axis system to construct the so-called substitution structure (*27*). The relative signs of the atom coordinates are the only additional parameters needed in this analysis, and they are assigned based on the ab initio cluster geometries (*16*). The substitution structures of the water hexamer isomers are compared to theoretical structures, which include vibrational averaging, in the top panel of Fig. 3. Zero-point vibrationally averaged structures are calculated using vibrational perturbation theory (VPT2) applied on the second-order Møller-Plesset perturbation theory (MP2) potential energy surface (*28*, *29*). In VPT2, the cubic and semidiagonal quartic force constants are computed by finite differentiation of the Hessian along the normal mode coordinates and are used to obtain the vibrationally averaged structure of a cluster at 0 K. A comparison of the vibrationally averaged and equilibrium theoretical geometries is shown in figs. S11 to S15. The direct Kraitchman analysis is sufficient to unambiguously identify the prism, cage, and book isomers.

However, the Kraitchman method is unreliable for quantitative analysis when atoms are close to the principal axes because of zero-point vibrational contributions to the moments of inertia that vary upon isotopic substitution. As detailed in the supplementary materials (tables S32, S35, and S38), each water hexamer isomer has two individual principal axis components that return unphysical values (out of the 18 total Cartesian coordinates that define the oxygen atom framework for a hexamer). An alternative means of analysis is to fit the structure directly to the complete set of isotopic rotational constants to determine the effective ground-state geometry, the so-called r_0 structure (*30*). The oxygen framework of 6 atoms is defined by $3N-6 = 12$ internal coordinates, and 21 experimental moments of inertia are available, so there is a sufficient number of degrees of freedom to fit the complete oxygen framework. Because the present experimental results carry only a limited amount of information on the positions of the hydrogen atoms, additional assumptions are required. The specific assumptions are not critical, but the lowest deviations of fits were obtained in the r_0 analysis assuming the experimental r_0 monomer geometry for each water unit (*31*). In addition, the orientations of the water molecules in the cluster are based on the vibrationally averaged theoretical structures in a way that keeps the angles of the strongest hydrogen bonds constant. The experimental r_0-structure oxygen atom frameworks are compared with the vibrationally averaged theoretical structures in the bottom panel of Fig. 3.

A detailed comparison of the experimental and theoretical O···O bond distances is given in Fig. 4. There is remarkable agreement between experiment and computations concerning the variation in O···O bond distances within each cluster isomer, allowing more confident, detailed insight into the hydrogen-bonding network for each isomer. The O···O distance is a known measure of hydrogen bond strength, and in the hexamer isomers these span the range 2.70 to 3.01 Å. This is a departure from the more uniform picture for the global minimum clusters up to the water pentamer in which symmetry enforces a single O···O distance, systematically decreasing with cluster size (*4*). The shortest O···O distance in the hexamers is shorter than the 2.76 Å separation of the pentamer, whereas the longest hexamer O···O distance exceeds the value found for the water dimer, 2.98 Å (*4*). The detachment energy of

Table 1. Experimental and calculated rotational constants for observed $(H_2{}^{16}O)_n$ clusters.

	Experimental	Theory	
		Equilibrium*	Vibrationally averaged
Hexamer cage			
A/MHz	2161.8762(8)	2249.2†	2172.4†
B/MHz	1129.3153(5)	1149.8	1129.2
C/MHz	1066.9627(5)	1097.9	1070.3
Hexamer prism			
A/MHz	1658.224(4)	1697.2	1644.5
B/MHz	1362.000(2)	1416.6	1373.6
C/MHz	1313.124(2)	1358.3	1312.5
Hexamer book			
A/MHz	1879.4748(8)	1898.3	1856.1
B/MHz	1063.9814(6)	1106.3	1080.5
C/MHz	775.0619(3)	809.5	788.7
Heptamer PR1			
A/MHz	1304.4355(3)	1333.7	1297.5
B/MHz	937.8844(6)	974.1	950.7
C/MHz	919.5236(6)	954.2	926.8
Nonamer D2dDD			
A/MHz	774.7442(7)	810.9	791.2
B/MHz	633.5403(6)	630.6	619.8
C/MHz	570.6460(5)	585.9	574.1

*From (*16*). †Distinct cage configurations are labeled according to the up(u) or down(d) position of the two free hydrogens of the edge (single donor-single acceptor) waters about the most immediate O-O-O plane. The calculated rotational constants are for the uu{1} structure, which is practically isoenergetic with the du{1} structure considered in (*16*) and is currently preferred on the basis of comparison between relative values of experimental and calculated dipole moment components and the deviations of the structural fits. See section 3.1 in the supplementary materials.

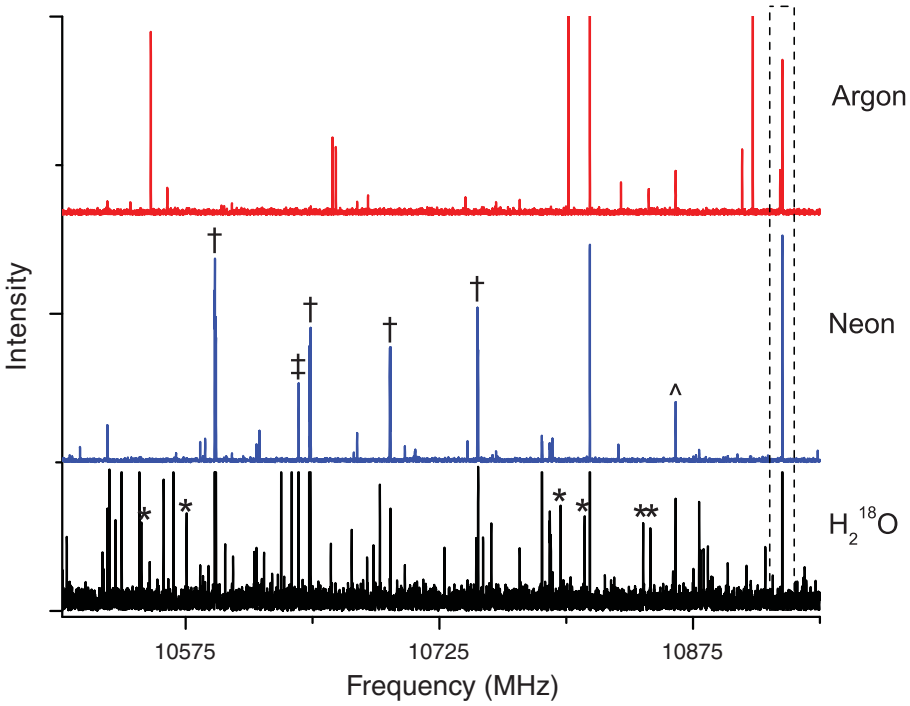


Fig. 2. The black trace shows a portion of the $H_2{}^{18}O$-enriched (15%) CP-FTMW spectrum obtained by averaging 550,000 FID signals. The offset blue trace shows the same region in the $H_2{}^{16}O$ cluster spectrum (average of 650,000 FID signals) measured with a neon backing gas. The offset red trace shows the same region (average of 200,000 FID signals) measured with an argon backing gas. The dashed box shows the 5_{05}-4_{04} rotational transition of the hexamer-cage isomer (intensity going off scale). The asterisks show the six individual $K_a = 0$ transitions resulting from isotopic substitution. The transitions marked with daggers indicate transitions assigned to the hexamer prism. The transition marked by the double dagger indicates a transition assigned to the hexamer book. The hexamer cage is the only isomer observed when argon is used as a backing gas, indicating that the cage is the lowest-energy isomer of the water hexamer. Similar data for the hexamer prism (fig. S4) and hexamer book (fig. S5) can be found in the supplementary materials. Finally, the caret indicates a transition belonging to the heptamer, whose presence remains when the backing gas is argon.

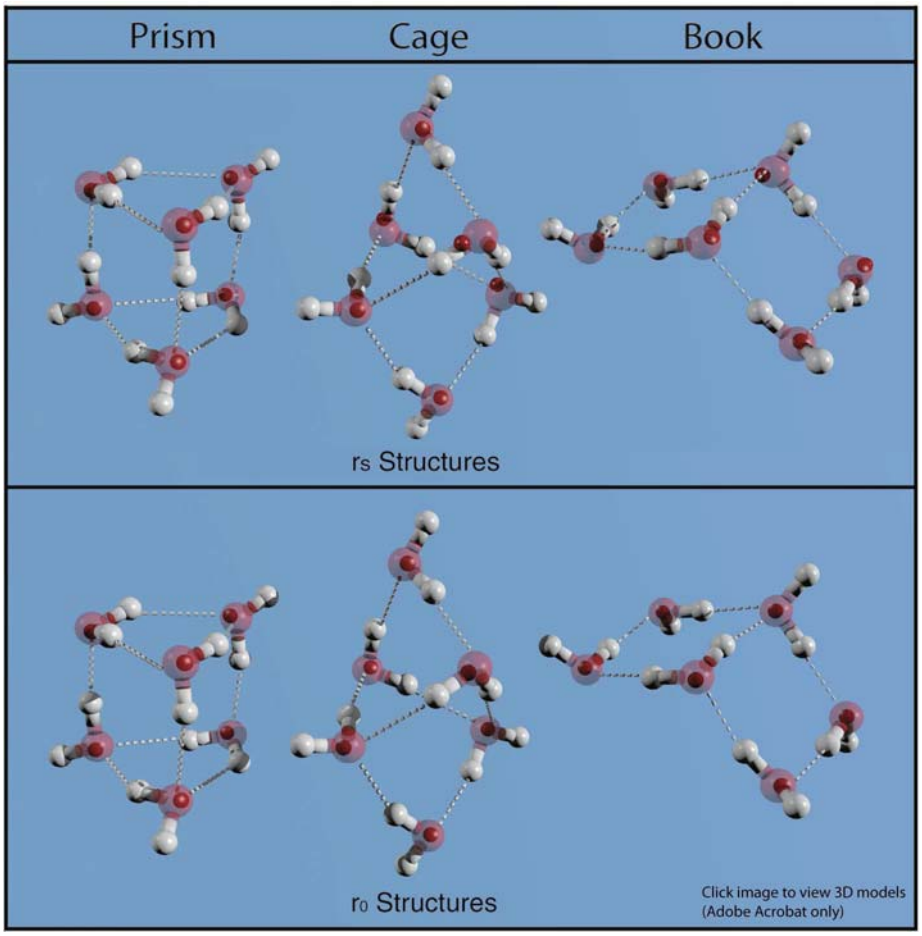


Fig. 3. The experimental structures of the water hexamer isomers determined using Kraitchman analysis (top, the substitution structure, r_s) and r_0 analysis (bottom) are compared with the vibrationally averaged ab initio structures. The full molecular structure rendering is the theoretical structure with lines drawn to denote the hydrogen-bonding network. The smaller dark red spheres are the experimental atom positions for the oxygen framework. Note the improved agreement in the oxygen atom positions for the r_0 analysis.

each water molecule in the cluster is also given in Fig. 4. This quantity was calculated with the density-fitted second-order Møller–Plesset perturbation theory with F12 explicit correlation factor employing a correlation consistent valence triple-zeta basis set (DF-MP2-F12/VTZ-F12) method (*32*), as detailed in the supplementary materials (section 1.2). The water detachment energies show large variation but correlate with the O···O distances and with the number of hydrogen bonds to each water monomer. This provides a consistent picture of the diversity of

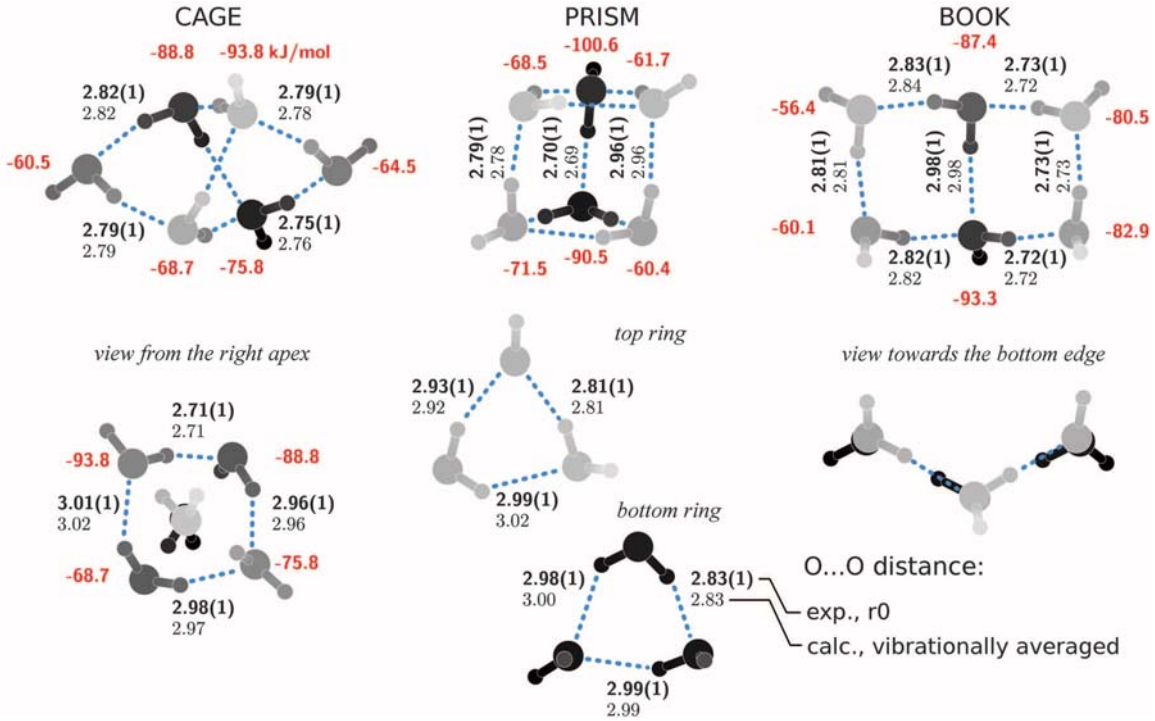


Fig. 4. The experimental r_0-analysis structures are shown for the three water hexamer isomers. The dashed lines indicate the hydrogen-bonding network. The experimental O···O bond distances, in Angstroms, are given in bold with the theoretical values from the vibrationally averaged structures below. The detachment energy for each water molecule in the cluster is given in red.

hydrogen bonding appearing at the hexamer cluster level and is a small-scale prelude to the known diversity in the structure of liquid water.

References and Notes

1. R. Bukowski, K. Szalewicz, G. C. Groenenboom, A. van der Avoird, *Science* **315**, 1249 (2007).
2. J. K. Gregory, D. C. Clary, *J. Phys. Chem.* **100**, 18014 (1996).
3. S. S. Xantheas, C. J. Burnham, R. J. Harrison, *J. Chem. Phys.* **116**, 1493 (2002).
4. K. Liu, J. D. Cruzan, R. J. Saykally, *Science* **271**, 929 (1996).
5. F. N. Keutsch, R. J. Saykally, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10533 (2001).
6. N. Pugliano, R. J. Saykally, *Science* **257**, 1937 (1992).
7. J. D. Cruzan *et al.*, *Science* **271**, 59 (1996).
8. K. Liu, M. G. Brown, J. D. Cruzan, R. J. Saykally, *Science* **271**, 62 (1996).
9. K. Liu *et al.*, *Nature* **381**, 501 (1996).
10. K. Nauta, R. E. Miller, *Science* **287**, 293 (2000).
11. C. Steinbach *et al.*, *Phys. Chem. Chem. Phys.* **6**, 3320 (2004).
12. R. N. Pribble, T. S. Zwier, *Science* **265**, 75 (1994).
13. C. J. Gruenloh *et al.*, *Science* **276**, 1678 (1997).
14. E. G. Diken, W. H. Robertson, M. A. Johnson, *J. Phys. Chem. A* **108**, 64 (2004).
15. G. G. Brown *et al.*, *Rev. Sci. Instrum.* **79**, 053103 (2008).
16. B. Temelso, K. A. Archer, G. C. Shields, *J. Phys. Chem. A* **115**, 12034 (2011).
17. U. Góra, R. Podeszwa, W. Cencek, K. Szalewicz, *J. Chem. Phys.* **135**, 224102 (2011).
18. G. Hincapié, N. Acelas, M. Castaño, J. David, A. Restrepo, *J. Phys. Chem. A* **114**, 7809 (2010).
19. Y. Chen, H. Li, *J. Phys. Chem. A* **114**, 11719 (2010).
20. D. M. Bates, G. S. Tschumper, *J. Phys. Chem. A* **113**, 3555 (2009).
21. E. E. Dahlke, R. M. Olson, H. R. Leverentz, D. G. Truhlar, *J. Phys. Chem. A* **112**, 3976 (2008).
22. B. Santra *et al.*, *J. Chem. Phys.* **129**, 194111 (2008).
23. K. Liu, M. G. Brown, R. J. Saykally, *J. Phys. Chem. A* **101**, 8995 (1997).
24. Y. Wang, J. M. Bowman, *J. Chem. Phys.* **134**, 154510 (2011).
25. E. Zwart, J. J. ter Meulen, W. Leo Meerts, L. H. Coudert, *J. Mol. Spec.* **147**, 27 (1991).
26. U. Erlekam, M. Frankowski, G. von Helden, G. Meijer, *Phys. Chem. Chem. Phys.* **9**, 3786 (2007).
27. J. Kraitchman, *Am. J. Phys.* **21**, 17 (1953).
28. V. Barone, *J. Chem. Phys.* **122**, 014108 (2005).
29. M. J. Frisch *et al.*, Gaussian 09 (Gaussian, Inc., Wellington, CT, 2009).
30. Z. Kisiel, *J. Mol. Spec.* **218**, 58 (2003).
31. F. C. De Lucia, P. Helminger, W. Gordy, H. Morgan, P. Staats, *Phys. Rev. A* **8**, 2785 (1973).
32. H. J. Werner, T. B. Adler, F. R. Manby, *J. Chem. Phys.* **126**, 164102 (2007).

Acknowledgments: This work was supported by U.S. National Science Foundation (NSF) grants CHE-0960074 and CHE-0848827; the Institute of Physics, Polish Academy of Sciences; and NSF grants CHE-0116435, CHE-0521063, and CHE-0849677 as part of the MERCURY high-performance computer consortium (www.mercuryconsortium.org). This research used the NSF TeraGrid resources provided by the Texas Advanced Computing Center (TACC) under grant TG-CHE090095 and resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under contract DE-AC02-05CH11231.

Supplementary Materials

www.sciencemag.org/cgi/content/full/336/6083/897/DC1
Materials and Methods
Figs. S1 to S15
Tables S1 to S46
References (*33*–*62*)

15 February 2012; accepted 23 March 2012
10.1126/science.1220574

Observation of ^{239}Pu Nuclear Magnetic Resonance

H. Yasuoka,[1,2] G. Koutroulakis,[1]* H. Chudo,[1,2] S. Richmond,[1] D. K. Veirs,[1] A. I. Smith,[1] E. D. Bauer,[1] J. D. Thompson,[1] G. D. Jarvinen,[1] D. L. Clark[1]

In principle, the spin-½ plutonium-239 (^{239}Pu) nucleus should be active in nuclear magnetic resonance spectroscopy. However, its signal has eluded detection for the past 50 years. Here, we report observation of a ^{239}Pu resonance from a solid sample of plutonium dioxide (PuO_2) subjected to a wide scan of external magnetic field values (3 to 8 tesla) at a temperature of 4 kelvin. By mapping the external field dependence of the measured resonance frequency, we determined the nuclear gyromagnetic ratio $^{239}\gamma_n(PuO_2)/2\pi$ to be 2.856 ± 0.001 megahertz per tesla (MHz/T). Assuming a free-ion value for the Pu^{4+} hyperfine coupling constant, we estimated a bare $^{239}\gamma_n/2\pi$ value of ~2.29 MHz/T, corresponding to a nuclear magnetic moment of $\mu_n \approx 0.15\mu_N$ (where μ_N is the nuclear magneton).

Since its discovery by Bloch and Purcell in 1946 (*1*, *2*), nuclear magnetic resonance (NMR) spectroscopy has evolved into an indispensable tool for experimental condensed matter physics, the determination of molecular structure, the study of molecular dynamics, and the characterization of atoms and molecules in the gaseous, liquid, and solid state. NMR is predicated on the fact that a nucleus with a nonzero spin angular momentum **I** possesses a magnetic

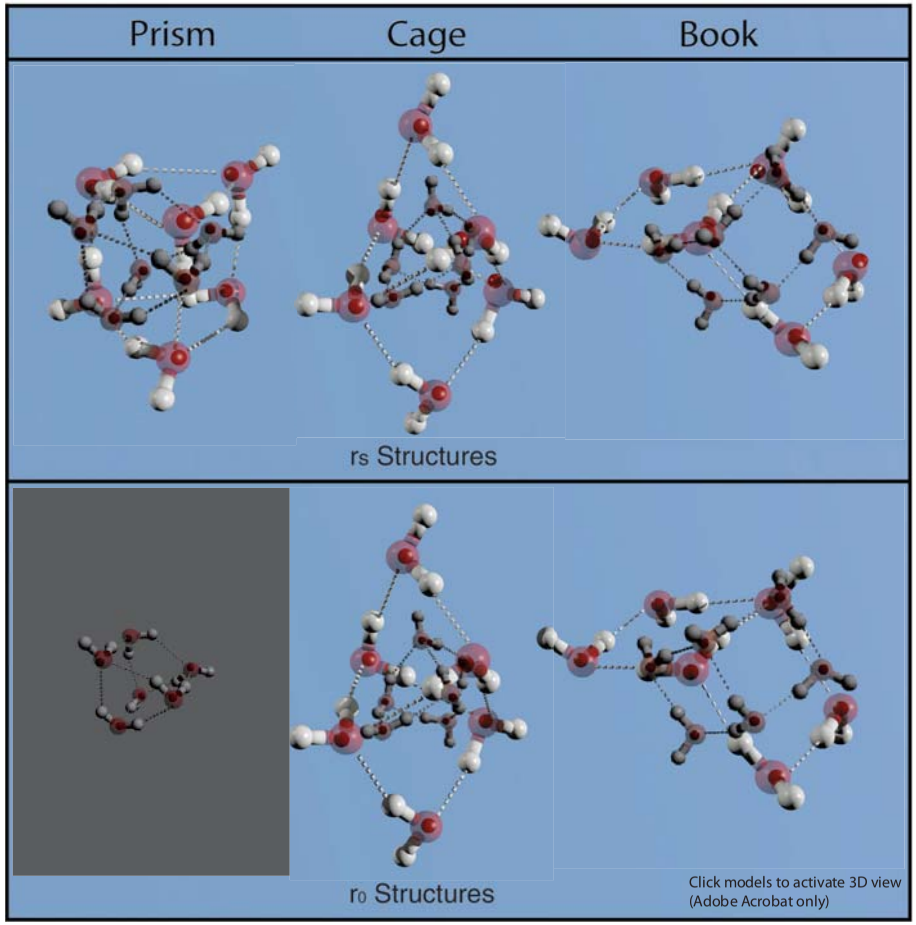


Fig. 3. The experimental structures of the water hexamer isomers determined using Kraitchman analysis (top, the substitution structure, r_s) and r_0 analysis (bottom) are compared with the vibrationally averaged ab initio structures. The full molecular structure rendering is the theoretical structure with lines drawn to denote the hydrogen-bonding network. The smaller dark red spheres are the experimental atom positions for the oxygen framework. Note the improved agreement in the oxygen atom positions for the r_0 analysis.

[Note to Acrobat users: If a warning message appears, select "Options" and "trust this host once"; then click on models to activate.]

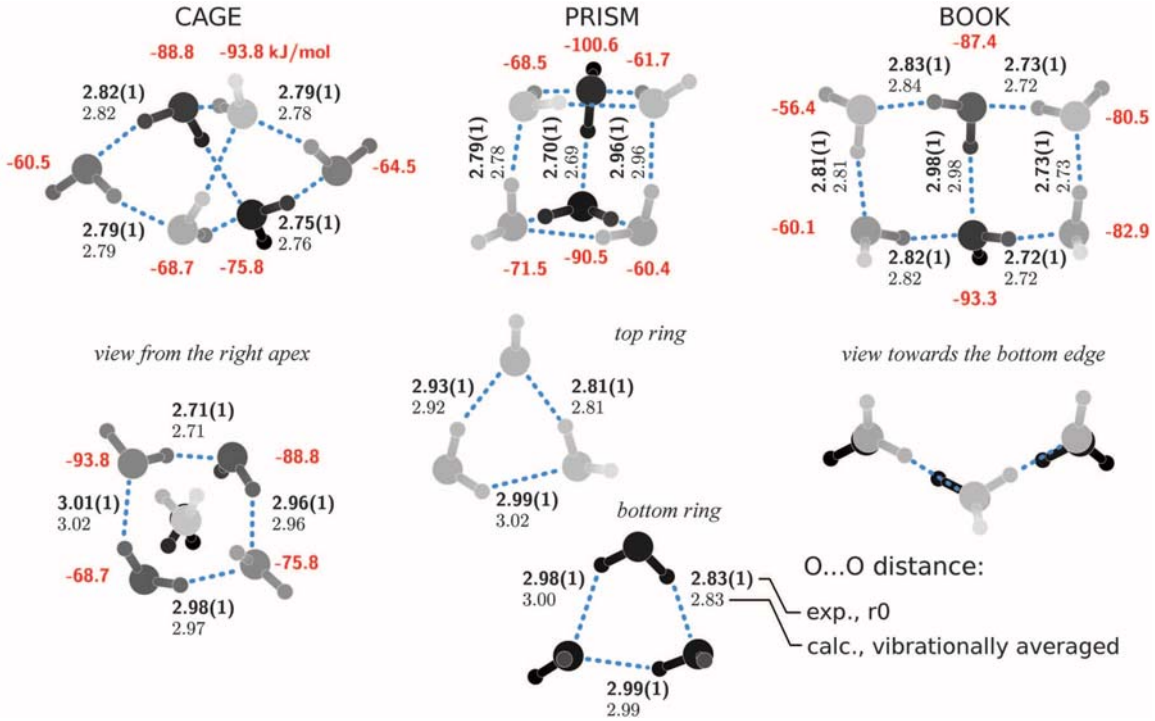


Fig. 4. The experimental r_0-analysis structures are shown for the three water hexamer isomers. The dashed lines indicate the hydrogen-bonding network. The experimental O···O bond distances, in Angstroms, are given in bold with the theoretical values from the vibrationally averaged structures below. The detachment energy for each water molecule in the cluster is given in red.

hydrogen bonding appearing at the hexamer cluster level and is a small-scale prelude to the known diversity in the structure of liquid water.

References and Notes

1. R. Bukowski, K. Szalewicz, G. C. Groenenboom, A. van der Avoird, *Science* **315**, 1249 (2007).
2. J. K. Gregory, D. C. Clary, *J. Phys. Chem.* **100**, 18014 (1996).
3. S. S. Xantheas, C. J. Burnham, R. J. Harrison, *J. Chem. Phys.* **116**, 1493 (2002).
4. K. Liu, J. D. Cruzan, R. J. Saykally, *Science* **271**, 929 (1996).
5. F. N. Keutsch, R. J. Saykally, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10533 (2001).
6. N. Pugliano, R. J. Saykally, *Science* **257**, 1937 (1992).
7. J. D. Cruzan *et al.*, *Science* **271**, 59 (1996).
8. K. Liu, M. G. Brown, J. D. Cruzan, R. J. Saykally, *Science* **271**, 62 (1996).
9. K. Liu *et al.*, *Nature* **381**, 501 (1996).
10. K. Nauta, R. E. Miller, *Science* **287**, 293 (2000).
11. C. Steinbach *et al.*, *Phys. Chem. Chem. Phys.* **6**, 3320 (2004).
12. R. N. Pribble, T. S. Zwier, *Science* **265**, 75 (1994).
13. C. J. Gruenloh *et al.*, *Science* **276**, 1678 (1997).
14. E. G. Diken, W. H. Robertson, M. A. Johnson, *J. Phys. Chem. A* **108**, 64 (2004).
15. G. G. Brown *et al.*, *Rev. Sci. Instrum.* **79**, 053103 (2008).
16. B. Temelso, K. A. Archer, G. C. Shields, *J. Phys. Chem. A* **115**, 12034 (2011).
17. U. Góra, R. Podeszwa, W. Cencek, K. Szalewicz, *J. Chem. Phys.* **135**, 224102 (2011).
18. G. Hincapié, N. Acelas, M. Castaño, J. David, A. Restrepo, *J. Phys. Chem. A* **114**, 7809 (2010).
19. Y. Chen, H. Li, *J. Phys. Chem. A* **114**, 11719 (2010).
20. D. M. Bates, G. S. Tschumper, *J. Phys. Chem. A* **113**, 3555 (2009).
21. E. E. Dahlke, R. M. Olson, H. R. Leverentz, D. G. Truhlar, *J. Phys. Chem. A* **112**, 3976 (2008).
22. B. Santra *et al.*, *J. Chem. Phys.* **129**, 194111 (2008).
23. K. Liu, M. G. Brown, R. J. Saykally, *J. Phys. Chem. A* **101**, 8995 (1997).
24. Y. Wang, J. M. Bowman, *J. Chem. Phys.* **134**, 154510 (2011).
25. E. Zwart, J. J. ter Meulen, W. Leo Meerts, L. H. Coudert, *J. Mol. Spec.* **147**, 27 (1991).
26. U. Erlekam, M. Frankowski, G. von Helden, G. Meijer, *Phys. Chem. Chem. Phys.* **9**, 3786 (2007).
27. J. Kraitchman, *Am. J. Phys.* **21**, 17 (1953).
28. V. Barone, *J. Chem. Phys.* **122**, 014108 (2005).
29. M. J. Frisch *et al.*, Gaussian 09 (Gaussian, Inc., Wellington, CT, 2009).
30. Z. Kisiel, *J. Mol. Spec.* **218**, 58 (2003).
31. F. C. De Lucia, P. Helminger, W. Gordy, H. Morgan, P. Staats, *Phys. Rev. A* **8**, 2785 (1973).
32. H. J. Werner, T. B. Adler, F. R. Manby, *J. Chem. Phys.* **126**, 164102 (2007).

Acknowledgments: This work was supported by U.S. National Science Foundation (NSF) grants CHE-0960074 and CHE-0848827; the Institute of Physics, Polish Academy of Sciences; and NSF grants CHE-0116435, CHE-0521063, and CHE-0849677 as part of the MERCURY high-performance computer consortium (www.mercuryconsortium.org). This research used the NSF TeraGrid resources provided by the Texas Advanced Computing Center (TACC) under grant TG-CHE090095 and resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under contract DE-AC02-05CH11231.

Supplementary Materials

www.sciencemag.org/cgi/content/full/336/6083/897/DC1
Materials and Methods
Figs. S1 to S15
Tables S1 to S46
References (*33*–*62*)

15 February 2012; accepted 23 March 2012
10.1126/science.1220574

Observation of ^{239}Pu Nuclear Magnetic Resonance

H. Yasuoka,[1,2] G. Koutroulakis,[1]* H. Chudo,[1,2] S. Richmond,[1] D. K. Veirs,[1] A. I. Smith,[1] E. D. Bauer,[1] J. D. Thompson,[1] G. D. Jarvinen,[1] D. L. Clark[1]

In principle, the spin-½ plutonium-239 (^{239}Pu) nucleus should be active in nuclear magnetic resonance spectroscopy. However, its signal has eluded detection for the past 50 years. Here, we report observation of a ^{239}Pu resonance from a solid sample of plutonium dioxide (PuO_2) subjected to a wide scan of external magnetic field values (3 to 8 tesla) at a temperature of 4 kelvin. By mapping the external field dependence of the measured resonance frequency, we determined the nuclear gyromagnetic ratio $^{239}\gamma_n(PuO_2)/2\pi$ to be 2.856 ± 0.001 megahertz per tesla (MHz/T). Assuming a free-ion value for the Pu^{4+} hyperfine coupling constant, we estimated a bare $^{239}\gamma_n/2\pi$ value of ~2.29 MHz/T, corresponding to a nuclear magnetic moment of $\mu_n \approx 0.15\mu_N$ (where μ_N is the nuclear magneton).

Since its discovery by Bloch and Purcell in 1946 (*1*, *2*), nuclear magnetic resonance (NMR) spectroscopy has evolved into an indispensable tool for experimental condensed matter physics, the determination of molecular structure, the study of molecular dynamics, and the characterization of atoms and molecules in the gaseous, liquid, and solid state. NMR is predicated on the fact that a nucleus with a nonzero spin angular momentum **I** possesses a magnetic

dipole moment equal to $\boldsymbol{\mu_n} = g_I \mu_N \mathbf{I} = \gamma_n \hbar \mathbf{I}$, where the nuclear spin is expressed in units of Planck's constant h divided by 2π ($\hbar$), μ_N is the nuclear magneton, g_I is the g-factor of the nucleus, and γ_n is the nuclear gyromagnetic ratio. In the presence of a magnetic field $\mathbf{B_0} \equiv B_0 \hat{z}$ (where $\hat{z}$ is the unit vector in the field's direction), the Zeeman interaction $H_z = -\boldsymbol{\mu_n} \cdot \mathbf{B_0}$ lifts the $(2I + 1)$-fold degeneracy of the nuclear spin ground state, resulting in $2I + 1$ equally spaced energy levels with $\Delta E \equiv h f_n = \gamma_n \hbar B_0$ (where f_n is frequency). In an NMR experiment, this energy ΔE is delivered to the nucleus by an appropriate radio frequency field transverse to $\mathbf{B_0}$, achieving a resonance governed by the relation $2\pi f_n = \gamma_n B_0$. From the latter equation, it is evident that γ_n effectively constitutes the fingerprint of a specific nucleus. In real materials, the nuclear spins sense an additional, internal magnetic field arising from their interaction with their electronic environment, as well as with one another. The static part of this internal field results in a change of the measured resonance frequency, and the resonance condition can be written as (Eq. 1)

$$2\pi f_n = \gamma_n B_0 (1 + K) \qquad (1)$$

The constant K (i.e., the NMR shift) represents a measure of the relative shift of the resonance frequency from that of the isolated nucleus and is equal to $A_{hf}\chi_0$, where A_{hf} is the hyperfine coupling constant, and χ_0 is the static and uniform spin susceptibility. The fluctuating part of the internal field, on the other hand, is responsible for the relaxation of nuclear spins.

From the approximately 90 nuclear isotopes that are known to have a nonzero magnetic moment, about 30 feature a nuclear spin of $I = ½$. The study of this class of nuclei is generally advantageous because their NMR properties are dictated solely by magnetic interactions, avoiding effects such as additional shifts or line broadening due to the nuclear quadrupole interaction (*3*). The most commonly leveraged spin-½ nucleus in this context is ^{1}H, the probing of which has had a profound impact on structural analysis and magnetic resonance imaging in physics, chemistry, and medicine. Plutonium-239 is the only spin-½ nucleus that has yet to be observed by NMR, in spite of more than 50 years of effort on a range of ^{239}Pu compounds (*4*).

In actinide science, in general, NMR studies have been limited in scope to nuclei associated with ligand atoms. The only exception is the application of ^{235}U NMR to UO_2 (*5*) and UF_6 (*6*): In the former case, a signal could be detected only in the compound's antiferromagnetic state (*5*). There have been extensive efforts to realize NMR on the actinide nuclei of various compounds because their electronic properties are governed predominantly by the actinide atom itself. The search for NMR in actinides has focused primarily on ^{239}Pu. As mentioned above, the spin-½ nucleus of ^{239}Pu with its sizable nuclear moment should, in principle, be an ideal candidate for NMR measurements in Pu-based compounds.

There are two main reasons why ^{239}Pu NMR has remained elusive. First, in atoms with unpaired f electrons (for example, lanthanides and actinides), an extremely strong hyperfine interaction between electron and nuclear spins gives rise to a large internal magnetic field (~100 T) at the nuclear site. As a consequence, the resonance frequency is shifted by several orders of magnitude, and the nuclear spin-lattice relaxation rate ($1/T_1$; where T_1 is the relaxation time) becomes exceedingly fast ($T_1 << 1$ μs), rendering any NMR measurement very challenging. Second, there is no accurate account of the ^{239}Pu nuclear moment μ_n, hence, neither is there a reliable account of the relevant value of γ_n. Various experimental probes (*7*–*10*) have aimed to determine μ_n indirectly, but the estimated values range between 0.11 and 0.39μ_N (or, equivalently, $^{239}\gamma_n/2\pi \approx 1.6$ to 6 MHz/T), making the possible parameter space for an NMR resonance quite extended.

To overcome the aforementioned challenge imposed by the excessively strong hyperfine interaction, we undertook experiments on PuO_2, in which the Pu ion is in a tetravalent state and has cubic local symmetry (see fig. S1). Although PuO_2 is somewhat nonstoichiometric, Pu ions in a fully oxidized sample should be Pu^{4+} ($5f^4$, $3H^4$) with a Γ_1 singlet ground state. With the first excited magnetic state Γ_4 found to be 123 meV higher in energy than the ground state (*11*), the Pu^{4+} ion should be nearly nonmagnetic at relatively low temperature. This feature is reflected in the observed temperature-independent magnetic susceptibility (*12*), unlike the strong Curie-Weiss–like temperature dependence of the susceptibility found in UO_2 and NpO_2 (*13*, *14*). As for the unknown γ_n value, we swept the external magnetic field at a constant frequency over the full range, spanning the minimum and the maximum expected values.

We performed our measurements on a high-purity PuO_2 powder sample, synthesized by nitric acid anion exchange followed by oxalate precipitation and calcination (*15*). The ^{239}Pu isotopic abundance was ~94%. The sample had an O/Pu ratio of 2.06, indicating a slight excess of oxygen atoms at interstitial sites. Nevertheless, a Rietveld analysis of the x-ray powder diffraction pattern confirmed that the sample was pure cubic PuO_2 without any secondary phases.

We used a standard pulsed NMR technique to search for the ^{239}Pu NMR signal (*15*). The first trial run consisted of a broad field sweep, from $B_0 = 3$ to 8 T, with a step size of $\Delta B_0 = 0.06$ T at a constant frequency $f = 16.51$ MHz and temperature $T = 4$ K. The deduced integrated spin-echo intensity as a function of B_0 is depicted in Fig. 1, where a clear signal is evident at $B_0 \approx 5.8$ T. Having detected this signal, we optimized the experimental NMR parameters and performed detailed field-sweep measurements. Specifically, we carefully traced the observed signal for different carrier frequencies f to derive the precise value of γ_n. A representative field-swept NMR spectrum for $f = 20.48$ MHz is plotted in Fig. 2A. Using a simple Gaussian fit to the NMR peak, we determined the resonance field value and the full width at half maximum (FWHM), which, for the case of the spectrum of Fig. 2A, are $B_{res} = 7.1716$ T and FWHM ≈ 0.035 T, respectively. A sketch illustrating the experimental procedure is shown in Fig. 2B. Mapping the evolution of B_{res} as a function of f, we constructed the graph of Fig. 3. The value of γ_n in PuO_2 is given by the slope of this graph, derived from a linear fit to the data as $^{239}\gamma_n(PuO_2)/2\pi = 2.856 \pm 0.001$ MHz/T.

We were unable to precisely measure T_1, but we can place a lower bound of $T_1 \sim 100$ s at $T = 4$ K (*15*). A long T_1 is expected at low temperatures, given the singlet ground state of Pu^{4+} in PuO_2. Thus, the observed slow nuclear relaxation further confirms that the detected NMR signal is really associated with Pu^{4+}.

The γ_n value of ^{239}Pu obtained here for the case of PuO_2 does not necessarily represent the

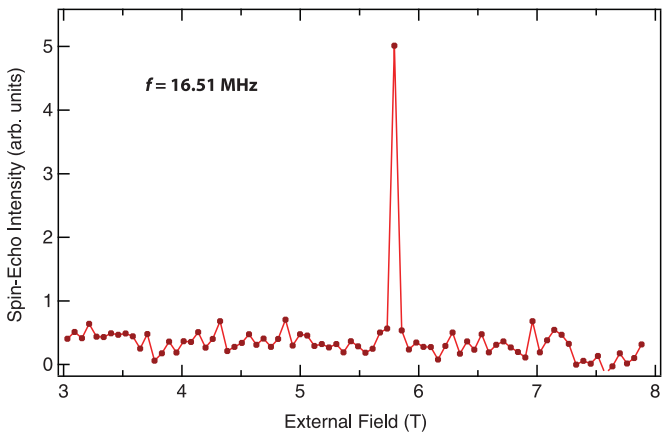


Fig. 1. Field-sweep ^{239}Pu NMR spectrum for $f = 16.51$ MHz and $T = 4$ K. Dots correspond to the integrated spin-echo intensity for each B_0 value. A signal is evident at $B_0 \approx 5.8$ T. The red solid line is a guide to the eye.

[1]Los Alamos National Laboratory (LANL), Los Alamos, NM 87545, USA. [2]Advanced Science Research Center, Japan Atomic Energy Agency, Tokai, Ibaraki 319-1195, Japan.

*To whom correspondence should be addressed. E-mail: gkoutrou@lanl.gov

bare γ_n of this nucleus. Although, as discussed earlier, the ground state of Pu^{4+} in PuO_2 is a nonmagnetic singlet, a Van Vleck paramagnetic contribution due to higher-energy magnetic states is expected even at low temperatures. This has been verified by magnetic susceptibility measurements of PuO_2, which afforded a temperature-independent susceptibility of $\chi_0 = 5.36 \times 10^{-4}$ emu/mol (emu, electromagnetic units) (*12*). The finite χ_0 translates into a shift for our NMR measurements. Therefore, to obtain the bare γ_n value of ^{239}Pu, the NMR shift must be estimated and accounted for, according to the resonance condition of Eq. 1. However, A_{hf} for the ^{239}Pu nucleus in PuO_2 is not known. In general, for free actinide ions, the hyperfine interaction is dominated by the term related to the orbital angular momentum of open-shell electrons (*16*). This term is largely dependent on the value of $\langle 1/r^3 \rangle$, where r is the distance between the nucleus and the respective interacting electron, and $\langle ... \rangle$ denotes an average over all unpaired electrons. The relevant calculated values of A_{hf} for the free actinide ions U^{4+}, Np^{4+}, and Pu^{4+} are 143.8, 180.3, and 258.3 T/μ_B, respectively. In solids, however, these free-ion values may decrease due to a reduction of $\langle 1/r^3 \rangle$. Nevertheless, NMR and neutron-scattering experiments in UO_2, which shares the same crystal structure with PuO_2, have deduced $A_{hf} = 145$ T/μ_B for the nucleus of U^{4+} (*5*, *17*). Surprisingly, this value is in excellent agreement with the calculated free-ion value, indicating that the quantity $\langle 1/r^3 \rangle$ remains nearly unchanged for U^{4+} in UO_2. Similarly, for the case of various Np-based compounds, Mössbauer spectroscopy offers a value of $A_{hf} = 192$ T/μ_B (*18*), which, again, is very close to the calculated free-ion value. We can thus infer that equating A_{hf} in actinide compounds with that of the relevant free ions constitutes a reasonable approximation. It should be noted, though, that this assumption is valid only for well-localized f-electron systems.

Based on these precedents, we assume the Pu^{4+} free-ion value of $A_{hf} = 258.3$ T/μ_B to extract K in PuO_2. Taking into account that $\chi_0 = 5.36 \times 10^{-4}$ emu/mol (9.594×10^{-4} μ_B/T) (*12*), we deduce $K \approx 24.8\%$. It is evident from Eq. 1 that the bare ^{239}Pu γ_n can then be derived as $^{239}\gamma_n = {}^{239}\gamma_n(PuO_2)/(1 + K) \approx 2\pi \cdot 2.29$ MHz/T (with an estimated precision of ~1%), which, in turn, corresponds to a nuclear magnetic moment of $\mu_n \approx 0.15\mu_N$.

Plutonium and its compounds show incredibly complex behavior, exhibiting a multiplicity of oxidation states and a pervasive tendency to form a range of nonstoichiometric phases in the solid state (*19*). Directly accessing the consequences of plutonium's 5f electrons at the atomic and structural unit-cell scales with NMR can provide a valuable tool for the study of plutonium solid-state physics, chemistry, and materials science. In view of the technological importance of plutonium compounds for nuclear fuels, power generation for interplanetary exploration, and long-term storage of nuclear waste, the ability to probe local structure, nuclear spin-relaxation processes, and electronic environment of the nucleus by ^{239}Pu NMR should prove especially powerful.

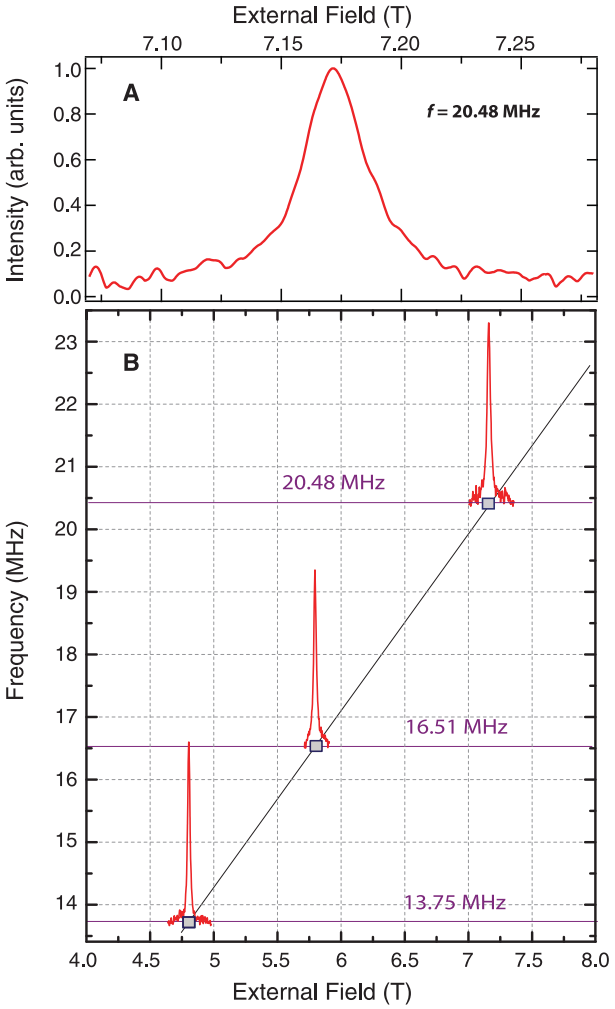


Fig. 2. (**A**) The ^{239}Pu NMR spectrum for $f = 20.48$ MHz and $T = 4$ K. A Gaussian fit gives $B_{res} = 7.1716$ T and FWHM ≈ 0.035 T. (**B**) Sketch of the experimental procedure: For several constant frequency values f (purple solid horizontal lines), ^{239}Pu field-swept NMR spectra were obtained (red solid lines) mapping the pairs of f, B_{res} values that satisfy the resonance condition $2\pi f = \gamma_n(PuO_2) \cdot B_{res}$ (solid squares). Dashed lines are guides to the eye.

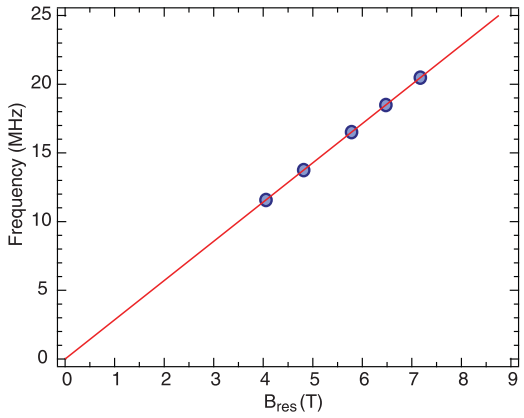


Fig. 3. Pairs of f, B_{res} values corresponding to the first moment of the recorded ^{239}Pu NMR spectra are indicated by the solid circles (error bars are within the symbol size). The relevant slope is determined by a linear fit to the data passing through the origin, giving a value of $^{239}\gamma_n(PuO_2)/2\pi = 2.856 \pm 0.001$ MHz/T.

References and Notes

1. F. Bloch, W. W. Hansen, M. Packard, *Phys. Rev.* **69**, 127 (1946).
2. E. M. Purcell, H. C. Torrey, R. V. Pound, *Phys. Rev.* **69**, 37 (1946).
3. A. Abragam, *Principles of Nuclear Magnetism* (Oxford Univ. Press, New York, 1961).
4. F. Y. Fradin, in *The Actinides: Electronic Structure and Related Properties*, A. J. Freeman, J. B. Darby Jr., Eds. (Academic Press, New York, 1974), vol. 1, p. 181.
5. K. Ikushima *et al.*, *J. Phys. Soc. Jpn.* **67**, 65 (1998).
6. H. Le Bail, C. Chachaty, P. Rigny, R. Bougon, *J. Physique Lett.* **44**, 1017 (1983).
7. J. Faust, R. Marrus, W. A. Nierenberg, *Phys. Lett.* **16**, 71 (1965).
8. R. J. Champeau, S. Gerstenkorn, *C. R. Hebd. Acad. Sci.* **251**, 352 (1960).
9. B. Bleaney, P. M. Llewellyn, M. H. L. Pryce, G. R. Hall, *Philos. Mag.* **45**, 991 (1954).
10. W. Kolbe, N. Edelstein, *Phys. Rev. B* **4**, 2869 (1971).
11. S. Kern *et al.*, *Phys. Rev. B* **59**, 104 (1999).
12. G. Raphael, R. Lallement, *Solid State Commun.* **6**, 383 (1968).
13. A. Arrott, J. E. Goldman, *Phys. Rev.* **108**, 948 (1957).

14. J. W. Ross, D. J. Lam, *J. Appl. Phys.* **38**, 1451 (1967).
15. Materials and methods are available as supplementary materials on *Science* Online.
16. B. D. Dunlap, G. M. Kalvius, in *The Actinides: Electronic Structure and Related Properties*, A. J. Freeman, J. B. Darby Jr., Eds. (Academic Press, New York, 1974), vol. 1, p. 237, and references therein.
17. J. Faber Jr., G. H. Lander, *Phys. Rev. B* **14**, 1151 (1976).
18. B. D. Dunlap, G. H. Lander, *Phys. Rev. Lett.* **33**, 1046 (1974).
19. D. L. Clark, S. S. Hecker, G. D. Jarvinen, M. P. Neu, in *The Chemistry of the Actinide and Transactinide Elements*, L. Morss, N. M. Edelstein, J. Fuger, Eds. (Springer, Berlin, 2006), vol. 2, p. 813.

Acknowledgments: H.Y. thanks the Glenn T. Seaborg Institute and the Los Alamos Laboratory Directed Research and Development (LDRD) program for financial support. This work was supported by the Chemical Sciences, Geosciences, and Biosciences Division and by the Materials Sciences and Engineering Division, Office of Basic Energy Sciences, U.S. Department of Energy (DOE). High-purity plutonium oxide was provided by the Surveillance and Monitoring Program at LANL supported by DOE Nuclear Materials Disposition EM-33. LANL is operated by Los Alamos National Security for the National Nuclear Security Administration of DOE under contract DE-AC52-06NA25396.

Supplementary Materials
www.sciencemag.org/cgi/content/full/336/6083/901/DC1
Materials and Methods
Figs. S1 to S4

20 February 2012; accepted 23 March 2012
10.1126/science.1220801

Conspecific Negative Density Dependence and Forest Diversity

Daniel J. Johnson,* Wesley T. Beaulieu, James D. Bever, Keith Clay

Conspecific negative density-dependent establishment, in which local abundance negatively affects establishment of conspecific seedlings through host-specific enemies, can influence species diversity of plant communities, but the generality of this process is not well understood. We tested the strength of density dependence using the United States Forest Service's Forest Inventory and Analysis database containing 151 species from more than 200,000 forest plots spanning 4,000,000 square kilometers. We found that most species experienced conspecific negative density dependence (CNDD), but there was little effect of heterospecific density. Additionally, abundant species exhibited weaker CNDD than rarer species, and species-rich regions exhibited stronger CNDD than species-poor regions. Collectively, our results provide evidence that CNDD is a pervasive mechanism driving diversity across a gradient from boreal to subtropical forests.

The factors that affect the distribution and abundance of species include abiotic influences, predators, parasites, competition, and mutualisms. A major mechanism proposed for the maintenance of diversity is conspecific negative density-dependent mortality, whereby proximity to adults of the same species reduces seedling survival through attack by host-specific enemies [also known as the Janzen-Connell hypothesis (JCH) (*1*, *2*)]. The JCH was proposed with particular reference to species-rich tropical forests. The relative importance of the JCH in influencing diversity in plant communities more widely has ramifications for accurate modeling and appropriate conservation and management.

Most studies of conspecific negative density dependence (CNDD) in plant communities have focused on a single site or a single species, [however, see (*3*–*5*) for comparisons of CNDD from multiple tropical forest sites]. Recent studies of individual tropical and subtropical forest sites suggest that forest composition is influenced by CNDD at the individual tree scale, which in turn influences the relative abundance of species at the community scale [(*6*–*8*), but see (*9*)]. In comparison, less is known about the efficacy of this mechanism in forests outside the tropical sites that were examined. Some temperate forest species display clear CNDD (*10*–*13*), and a comparative analysis of 13 temperate and tropical forest sites suggests that the proportion of species experiencing CNDD is not related to latitude (*10*). Despite the potential importance of CNDD in forest dynamics, we have a limited understanding of its role across forest communities.

Assessing CNDD at broader spatial scales is necessary for a more comprehensive understanding of the importance of CNDD to structuring forest communities. Here, we analyzed the U.S. Forest Service Forest Inventory and Analysis (FIA) database (http://fia.fs.fed.us/) (*14*). The volume and scope of the FIA data provided an opportunity to explore ecological patterns at wider scales than before. Plots were located from the Canadian border south to Florida and from the Atlantic coast west to the 100th meridian, approximately the center of the continental United States. We analyzed data on tree composition from fully forested plots (circular area of 168.33 m^2) and seedling establishment from nested plots (13.50 m^2) (fig. S1). FIA defines stems larger than 12.7 cm in diameter at breast height (dbh) as trees, whereas individuals smaller than 2.54-cm dbh and taller than 30.5 cm (12.24 cm for conifer species) are defined as seedlings. In total, we analyzed data from 207,444 paired tree and seedling plots, which contained 1,334,347 trees and 1,709,314 seedlings representing 151 species. Data were aggregated into 2° latitude-by-longitude regional cells to examine regional patterns of CNDD (fig. S2). Species richness per cell ranged from 18 to 119 (mean ± SE: 72.97 ± 2.26), and species richness per plot ranged from 1 to 21 (4.2 ± 0.005).

As CNDD is manifested by a negative relation between the abundance of conspecific trees versus seedlings, we estimated the strength of CNDD by regressing seedling abundance (S) on tree abundance (T) using generalized linear regression. The specific equation used to estimate the strength of CNDD by maximum likelihood was the negative exponential function $S = ae^{bT}$, where a is the y intercept and b controls the inflection of the curve (fig. S3). This provides a single parameter, b, to represent the strength of CNDD, with larger, more negative values of b indicating stronger CNDD. We repeated this analysis using basal area, a proxy for biomass, as an alternative measure of tree abundance.

In separate analyses, we fit the model for both conspecific and heterospecific trees for each species in each regional cell. We restricted our analyses to species that occurred in a minimum of 30 plots per cell where at least one plot had co-occurrence of trees with conspecific seedlings. Regional cells lacking co-occurrence of trees and seedlings represent the strongest cases of CNDD, making our analyses conservative. Moreover, the FIA definition of a seedling excludes most species' first-year seedlings and, therefore, does not capture early mortality, also making our analyses conservative. We examined the effect of heterospecific density to determine the role of density-dependent effects that were independent of conspecific abundance (for instance, shading by canopy trees or competition for soil nutrients).

Species by regional cell estimates of conspecific density dependence were significantly more negative and variable than heterospecific density dependence (Fig. 1, A and B), indicating much stronger and more varied effects of conspecific tree density on seedling establishment than heterospecific tree density. Analyses of basal area produced qualitatively similar results (Fig. 1, C and D). The stronger negative effect of conspecifics on seedling density occurred, despite the fact that seed density is expected to be higher near conspecific trees (*15*). Moreover, the reduction in conspecific seedling density cannot be totally explained by shading or soil-nutrient depletion because we did not find equivalent responses to heterospecific tree density or basal area (Fig. 1).

Conspecific negative density dependence of seedling establishment could provide a strong local filter favoring species diversity. For example, species experiencing strong CNDD will have fewer opportunities to establish, potentially reducing their relative abundance. Conversely,

Department of Biology, Indiana University, Bloomington, IN 47405, USA.

*To whom correspondence should be addressed. E-mail: dj4@indiana.edu

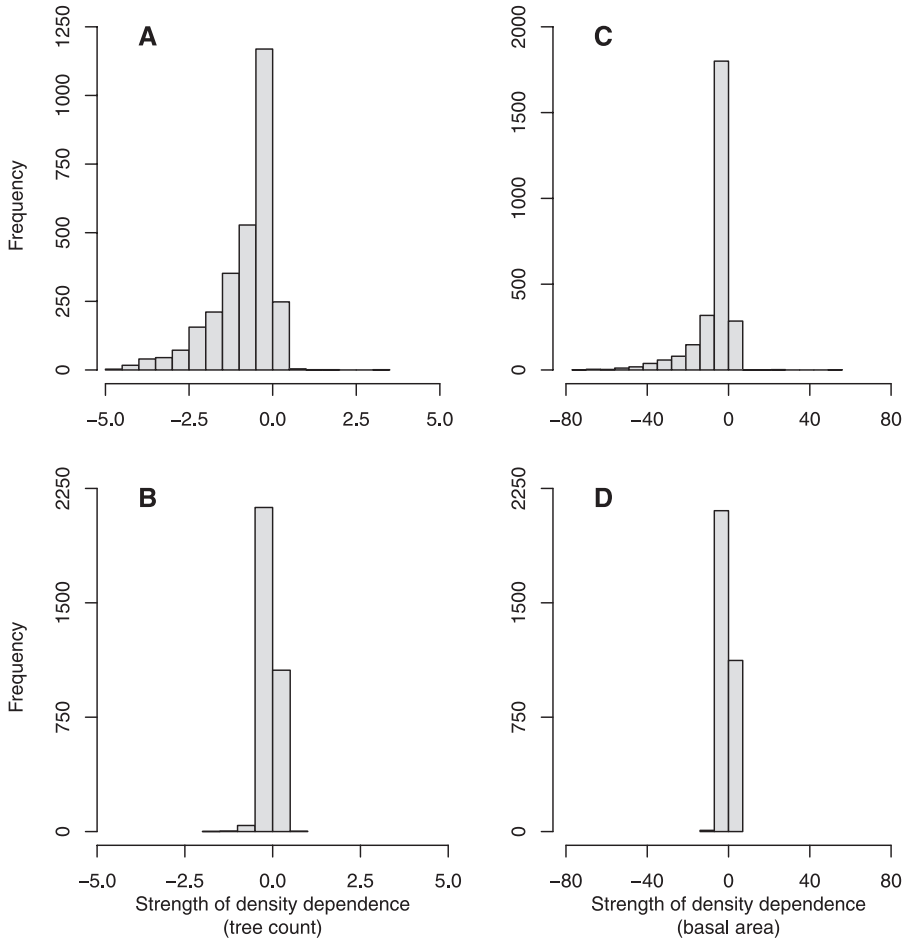


Fig. 1. Seedlings are significantly more likely to experience negative effects on establishment from conspecific neighbors than from heterospecific neighbors. The strength of density-dependent establishment was estimated for each species by regional cell combination on seedlings by (**A**) conspecific tree count (2848 combinations) and (**B**) heterospecific tree count (3229 combinations); these distributions were significantly different (Wilcoxon rank sum test, $W = 790{,}869$, $P < 2.2 \times 10^{-16}$). We performed the same analysis with (**C**) conspecific basal area (2768 combinations) and (**D**) heterospecific basal area (3234 combinations) and found that these distributions followed the same patterns and are significantly different (Wilcoxon rank sum test, $W = 1{,}362{,}121$, $P < 2.2 \times 10^{-16}$).

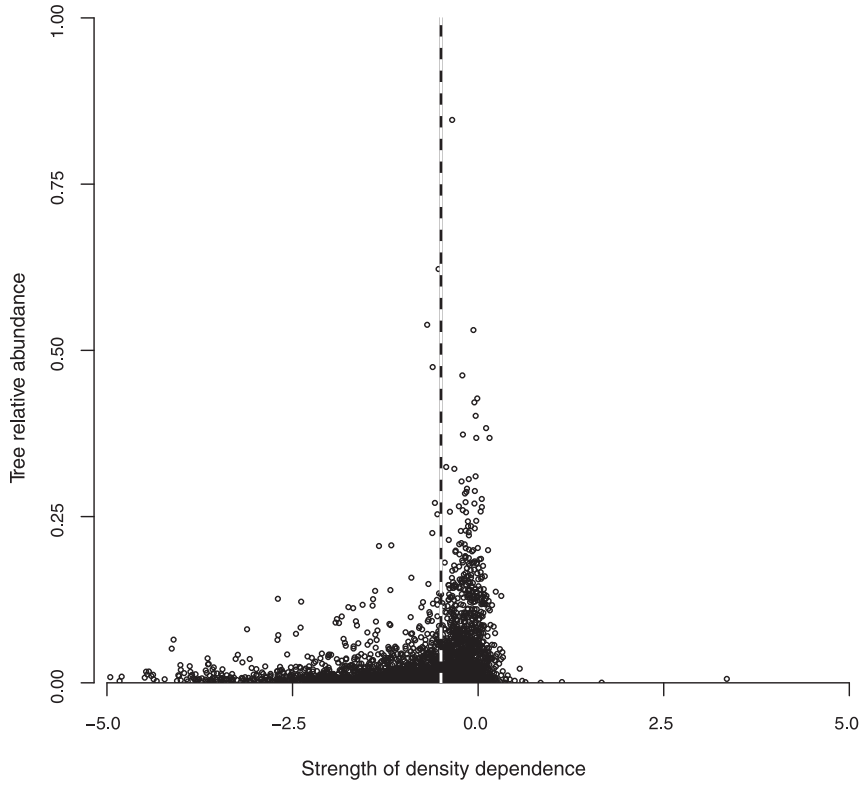


Fig. 2. Species experiencing strong conspecific density dependence have reduced relative abundance. The strength of conspecific density dependence had a significant positive correlation with species relative abundance (Spearman's rank correlation, $\rho = 0.3978$, $P < 2.2 \times 10^{-16}$; $N = 2848$ species by regional cell combinations). The dashed line represents the median of the strength of conspecific density dependence. Heterospecific density dependence (fig. S12) had a weaker trend in the opposite direction than conspecific density dependence (Spearman's rank correlation, $\rho = -0.1278$, $P = 3.121 \times 10^{-13}$; $N = 3229$ species by regional cell combinations). See fig. S5 for the corresponding conspecific basal area relation.

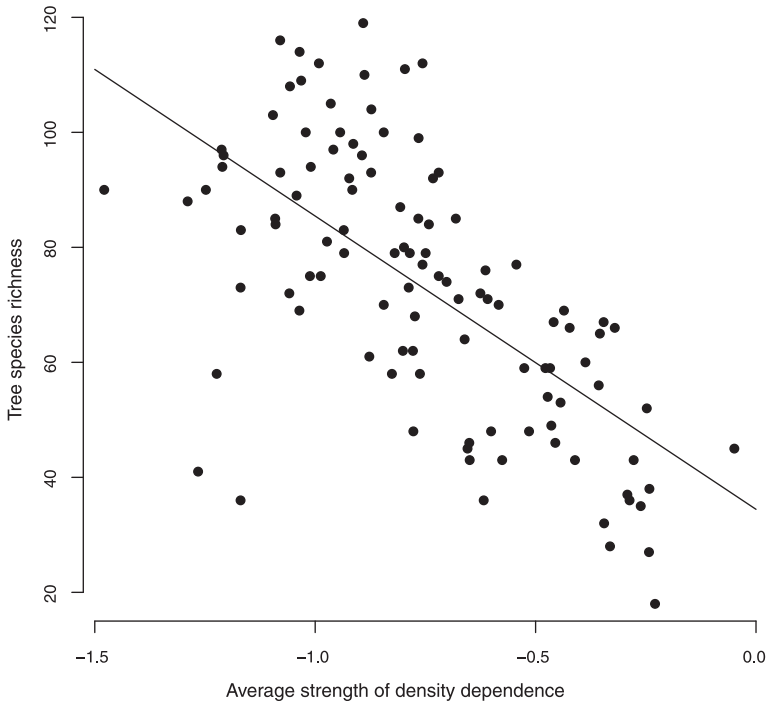


Fig. 3. Regional species richness is predicted by the strength of negative density dependence. The regional strength of conspecific density dependence, the arithmetic mean of all density dependence estimates in each regional cell, had a significant negative linear relation with species richness (correlation coefficient $r^2 = 0.42$, $F_{1,106} = 75.4$, $P = 5 \times 10^{-14}$; $N = 108$ regional cells) such that species richness tended to increase when negative density dependence was strongest. There was no significant relation between regional strength of heterospecific density dependence and species richness ($r^2 = 0.004$, $F_{1,106} = 0.5$, $P = 0.48$, $N = 108$ regional cells). See fig. S6 for the corresponding conspecific basal area relation.

species that exhibit weak CNDD could establish near conspecifics, thus allowing them to increase in relative abundance. We tested the prediction that the strength of density dependence influences relative tree abundance and found that species undergoing strong CNDD had reduced relative abundance in that regional cell relative to species experiencing weak CNDD (Fig. 2). We examined the intraspecific pattern of CNDD across the range of all species and found that the majority of species behave similarly, where individual species experiencing stronger CNDD are less common regionally (fig. S4). In contrast, heterospecific density dependence had a far weaker relationship with regional relative abundance, as expected with little variation in the strength of heterospecific density dependence. A gradient in the strength of CNDD and mainly weak heterospecific density dependence is consistent with observations at the Forest Dynamics Plot on Barro Colorado Island, Panama (*7*, *8*, *16*), and in greenhouse trials of abundant temperate species (*11*). Whereas our data represent one slice in time of longer-term community dynamics (*17*), our results conform to a recent theory (*8*, *18*, *19*) that predicts that the equilibrium relative abundance of a species will covary with the strength of conspecific density dependence.

The average strength of conspecific density dependence was strongly correlated with tree species richness at the regional cell level (Fig. 3). This finding suggests that local biotic interactions are underlying regional species richness, in contrast to the prevailing explanations of the gradient in tree-species richness in North America, which focus primarily on physical aspects of the environment [such as temperature and precipitation (*20*)]. Species-level differences are discounted

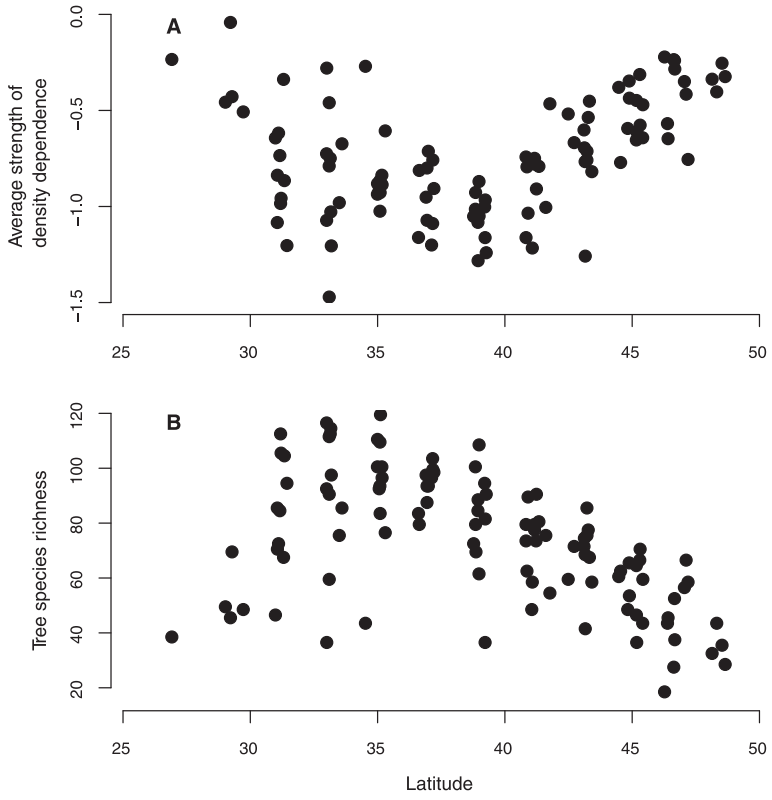


Fig. 4. The strength of conspecific density dependence becomes stronger with decreasing latitude. (**A**) The regional strength of conspecific density dependence, the arithmetic mean of all conspecific density dependence estimates in a 2° latitude-by-longitude regional cell, had a significant relation with increasing latitude (Spearman's rank correlation, $\rho = 0.3$, $P = 0.0003$; $N = 108$) (**B**). The pattern of tree species richness in the eastern United States does not follow a linear latitudinal gradient south of ~30° latitude because of reduced species richness resulting from the peninsular effect (Florida), reduced topographic heterogeneity, and arid southern Texas. See fig. S7 for the corresponding basal area relationship.

by neutral theory as unnecessary to describe the patterns of species abundances (*21*). Additionally, a recent statistical analysis has called into question the necessity of local interactions to describe patterns of diversity (*22*). Our results run counter to these arguments, as we found support for regional species richness patterns being driven by local species-specific ecological interactions and a local mechanism to explain variation in regional species richness.

It is possible that the patterns found here were generated by mechanisms unrelated to conspecific density dependence that could create spatial separation of adults and conspecific seedlings [e.g., timber harvesting, succession, the mass effect (*23*)]. For example, recruitment differences between early successional and late successional species could imitate patterns of CNDD in forests. To test whether CNDD varies with forest age, we reanalyzed the data set by stratifying the data into early (0 to 39 years), middle (40 to 79 years), and later (80+ years) successional forests. The patterns of CNDD were robust and consistent between age classes, indicating that our results are not contingent on successional dynamics or indirectly on timber harvesting, which has the effect of setting back forest age (figs. S8 to S11).

Janzen (*1*) and Connell (*2*) originally hypothesized that CNDD generated by host-specific seed predators could help maintain the high species richness in tropical forests. We found that CNDD is a strong mechanism maintaining species richness in eastern U.S. forests, but CNDD may also explain the latitudinal gradient in species richness if CNDD becomes stronger with decreasing latitude. We tested this hypothesis in eastern North America, where there is a latitudinal gradient of tree species richness that peaks in the southern Appalachian region (*20*). We found evidence that CNDD could maintain this gradient in tree species richness, as the average regional strength of CNDD was significantly negatively correlated with latitude, ranging from boreal to subtropical forests (Fig. 4). Our results suggest that the strength of CNDD would increase with decreasing latitude into species-rich tropical forests.

Our analyses of the FIA database provide robust evidence that CNDD is pervasive in forest communities and can significantly affect species relative abundance and species richness within and between forests. Further, our results show that species-specific processes acting on seedlings translate into patterns in the abundance and diversity of trees. Several potential interactions could generate CNDD, including intraspecific competition, autotoxicity, seed predators, and soil pathogens. Much research has demonstrated that the soil microbial community can drive CNDD in multiple plant communities, including tropical forests, temperate forests, grasslands, and sand dunes (*18*, *24*). In particular, two studies measuring soil community feedbacks, presumably driven by soil-borne pathogens, have identified a positive relation between strength of CNDD in the greenhouse and relative abundance in the field (*8*, *25*).

Local interactions have previously been considered a local filter on species diversity, but our findings indicate that local interactions feed back to regional species richness and abundance. Further, the prevalence of CNDD across many forest types and diverse species indicates the pervasive importance of these interactions. Our results show that CNDD is a general mechanism structuring forest communities across a wide gradient of forest types and can maintain the latitudinal gradient of tree species richness.

References and Notes

1. D. H. Janzen, *Am. Nat.* **104**, 501 (1970).
2. H. Connell, in *Dynamics of Populations*, P. J. G. den Boer, G. R. Gradwell, Eds. (Center for Agricultural Publishing and Documentation, Wageningen, Netherlands, 1971), pp. 298–312.
3. C. Wills *et al.*, *Science* **311**, 527 (2006).
4. S. J. Wright, *Oecologia* **130**, 1 (2002).
5. J. Terborgh, *Am. Nat.* **179**, 303 (2012).
6. L. Chen *et al.*, *Ecol. Lett.* **13**, 695 (2010).
7. L. S. Comita, H. C. Muller-Landau, S. Aguilar, S. P. Hubbell, *Science* **329**, 330 (2010).
8. S. A. Mangan *et al.*, *Nature* **466**, 752 (2010).
9. R. K. Kobe, C. F. Vriesendorp, *Ecol. Lett.* **14**, 503 (2011).
10. J. Hille Ris Lambers, J. S. Clark, B. Beckage, *Nature* **417**, 732 (2002).
11. S. McCarthy-Neumann, R. K. Kobe, *J. Ecol.* **98**, 408 (2010).
12. T. Nakashizuka, *Trends Ecol. Evol.* **16**, 205 (2001).
13. A. Packer, K. Clay, *Nature* **404**, 278 (2000).
14. Materials and methods are available as supplementary materials on *Science* Online.
15. J. S. Clark, M. Silman, R. Kern, E. Macklin, J. Hille Ris Lambers, *Ecology* **80**, 1475 (1999).
16. K. E. Harms, S. J. Wright, O. Calderón, A. Hernández, E. A. Herre, *Nature* **404**, 493 (2000).
17. K. Clay *et al.*, in *Infectious Disease Ecology: The Effects of Ecosystems on Disease and of Disease on Ecosystems*, R. S. Ostfeld, F. Keesing, V. T. Eviner, Eds. (Princeton Univ. Press, Princeton, NJ, 2008), pp. 145–178.
18. J. D. Bever, *New Phytol.* **157**, 465 (2003).
19. R. A. Chisholm, H. C. Muller-Landau, *Theor. Ecol.* **4**, 241 (2011).
20. D. J. Currie, V. Paquin, *Nature* **329**, 326 (1987).
21. S. P. Hubbell, *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, Princeton, NJ, 2001).
22. N. J. B. Kraft *et al.*, *Science* **333**, 1755 (2011).
23. A. Shmida, M. V. Wilson, *J. Biogeogr.* **12**, 1 (1985).
24. A. Kulmatiski, K. H. Beard, J. R. Stevens, S. M. Cobbold, *Ecol. Lett.* **11**, 980 (2008).
25. J. N. Klironomos, *Nature* **417**, 67 (2002).

Acknowledgments: We acknowledge contributions with database consulting from N. Long from the Indiana Univ. Research Database Complex, as well as statistical consulting from S. Dickinson and E. Hernandez at Indiana Statistical Consulting Center and C. Huang at the Indiana Univ. Department of Statistics. Helpful comments were provided by two anonymous reviewers, D. Civitello, C. Her, C. Lively, S. Mangan, S. McMahon, and S. Neumann. We also thank all of the FIA employees that collected and compiled the data that made these analyses possible. The data used for these analyses are publicly available from www.fia.fs.fed.us/tools-data/ under FIA Data Mart. See the supplementary materials for specific dates and conditions applied to the data set.

Supplementary Materials

www.sciencemag.org/cgi/content/full/336/6083/904/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S12
Table S1
References (*26*–*28*)

8 February 2012; accepted 19 April 2012
10.1126/science.1220269

Randomized Government Safety Inspections Reduce Worker Injuries with No Detectable Job Loss

David I. Levine,[1] Michael W. Toffel,[2]* Matthew S. Johnson[3]

Controversy surrounds occupational health and safety regulators, with some observers claiming that workplace regulations damage firms' competitiveness and destroy jobs and others arguing that they make workplaces safer at little cost to employers and employees. We analyzed a natural field experiment to examine how workplace safety inspections affected injury rates and other outcomes. We compared 409 randomly inspected establishments in California with 409 matched-control establishments that were eligible, but not chosen, for inspection. Compared with controls, randomly inspected employers experienced a 9.4% decline in injury rates (95% confidence interval = −0.177 to −0.021) and a 26% reduction in injury cost (95% confidence interval = −0.513 to −0.083). We find no evidence that these improvements came at the expense of employment, sales, credit ratings, or firm survival.

The U.S. Occupational Safety and Health Administration (OSHA) is one of the most controversial regulatory agencies in the United States. Some evidence indicates that OSHA penalties deter injuries (*1*), and OSHA supporters argue that inspections save lives at low cost to employers and employees and that additional regulation would reduce tens of thousands of occupational illnesses and hundreds of worker fatalities (*2*, *3*). At the same time, critics fear that OSHA destroys jobs without meaningfully improving workplace safety (*4*, *5*) and have urged the agency to shift its emphasis from worksite inspections to voluntary safety programs (*6*). Even if inspections do improve workplace safety, they might not be socially efficient if

[1]Haas School of Business, University of California, Berkeley, Berkeley, CA 94720, USA. [2]Harvard Business School, Boston, MA 02163, USA. [3]Department of Economics, Boston University, Boston, MA 02215, USA.

*To whom correspondence should be addressed. E-mail: mtoffel@hbs.edu

the cost of remediating hazards outweighs the benefits. The economic theory of perfectly competitive labor markets (with full information, perfect mobility of labor, and so forth) implies that remediating hazards will cause wages to decline so much that employees on average do not benefit from the increase in safety (*7*). If product markets are also perfectly competitive or if wages are sticky, then many inspected firms will either go out of business or at least suffer lower sales, lower employment, and worse credit ratings (*8*).

The debate has persisted in part because prior research has yielded widely varying results. For example, some studies find that OSHA inspections have little or no correlation with subsequent workplace injury rates (*9*–*11*), whereas others find that OSHA inspections correlate with a decline in injury rates (*1*, *12*–*14*). Similarly, workplace-safety inspections correlate with lower productivity in some studies (*15*) but not in others (*16*).

These widely varying results may be due in part to the substantial challenges of measuring the causal effect of OSHA inspections. One challenge arises because most OSHA inspections target workplaces with recent accidents or safety complaints, and these workplaces typically have a combination of ongoing safety problems and a random event ("bad luck") that year. Thus, a cross-sectional analysis revealing a positive correlation between inspections and subsequent injuries does not imply that OSHA inspections cause injuries; it could just be due to ongoing safety problems that spurred the inspection. At the same time, because the random element that contributes to an accident or complaint is temporary, injuries rates often revert to prior levels (*17*), and so inspections often precede a decline in injuries without necessarily causing the improvement, potentially biasing a panel data analysis of targeted inspections.

In addition, most previous studies of the effects of inspections analyze data from logs of workplace injuries that OSHA requires companies to maintain at each workplace. OSHA mandates better recordkeeping when its inspections find incomplete logs, which can erroneously make it appear as if inspections cause higher injury rates. For example, the injury rates reported by very large manufacturing plants more than doubled in the late 1980s after OSHA imposed multimillion dollar fines on a few such plants for poor recordkeeping (*18*).

Fortunately for evaluation purposes, California's Division of Occupational Safety and Health (Cal/OSHA) randomly selected workplaces in high-injury industries for inspections in 1996 to 2006 (*19*). By focusing on these inspections, we simulated a randomized controlled trial that can provide unbiased estimates of the effects of OSHA inspections. To do so, we matched on observables to construct a control group of very similar facilities that were eligible for randomized inspections but not selected.

In addition, we analyzed injury data from the workers' compensation system. Unlike OSHA-mandated logs, workers' compensation data are less likely to be affected by improved recordkeeping after OSHA inspections. Finally, because injuries are not the only outcome that might be affected by OSHA inspections, we also analyzed employment, company survival, and compensation to look for unintended harms from inspections.

The starting point of our analysis was to understand how Cal/OSHA selected establishments for randomized inspections. In each year of our study period (1996–2006), Cal/OSHA identified a list of industries with high injury rates—typically based on data from the U.S. Bureau of Labor Statistics (*19*)—for that year's randomized inspections. For each of these industries, Cal/OSHA used Dun & Bradstreet and other sources to compile a list of establishments with 10 or more employees, then randomly selected a subset of each list. These subsets were then sent to the appropriate northern or southern district managers (each district covers roughly half the state). Within each district, inspectors attempted to inspect all of the randomly chosen establishments, although managers could prioritize on the basis of factors such as avoiding industries they felt were not as dangerous and skipping workplaces that had had an OSHA inspection in the prior 2 years. Our procedure for choosing a sample adjusts for these factors. Specifically, we found controls in the same industry, and we dropped all treatments and potential controls that had had inspections in the prior 2 years.

We obtained data on these inspections from U.S. OSHA's Integrated Management Information System (IMIS). We obtained annual establishment-level data on payroll and on the number and value of workers' compensation claims from the Uniform Statistical Reporting Plan database of the Workers' Compensation Insurance Rating Board (WCIRB). For all California establishments tracked by Dun & Bradstreet, we obtained annual establishment-level data on company names, addresses, whether the establishment was a standalone firm (not a branch or subsidiary), Standard Industrial Classification (SIC) and North American Industry Classification System (NAICS) industry codes, sales, and employment from the National Establishment Time-Series (NETS) database.

We began constructing our analysis sample by identifying in OSHA's IMIS database the 1752 establishments at which Cal/OSHA had attempted a random inspection at least once during our sample period. Because injury data from workers' compensation systems are available primarily at the company level, we restricted our analysis to single-establishment firms. Because Cal/OSHA performed random inspections only at establishments with at least 10 employees, we included only establishments with at least 10 employees in the random inspection year or either of the two preceding years. The pipeline of how we linked these treatments and a set of potential controls to the several data sets and then restricted potential controls to resemble treatments by requiring them to be in the same industry and the same region of California, to be classified as a single-establishment firm, to have 10 or more employees, and so forth is shown in table S1. When more than one potential control matched the industry and region of a particular treatment, we selected the one with the most similar number of employees.

This matching process resulted in a matched sample of 409 pairs of single-establishment firms, whose industry distribution is reported in table S2. At 7% of the treatment establishments in our sample, Cal/OSHA did not carry out the inspection, typically because the inspector could not find the establishment, the establishment had gone out of business, or the inspector determined that the establishment was not eligible for a random inspection after all (for example, if the inspector found out the establishment had fewer than 10 employees). As we could not filter the control sample on these criteria, we included as treatments all establishments in which Cal/OSHA had attempted an inspection. Thus, our estimates measure the causal effect of an attempted inspection and might slightly underestimate the causal effect of the inspections that actually occurred. However, as the vast majority of the attempts were successful, we usually simplify our language by dropping the qualifier "attempted" and referring to our estimates as the causal effect of inspections.

To reduce the effect of very large outliers, we top-coded our measures of injury count (the annual number of workers' compensation claims) and injury cost (the annual value of workers' compensation claims) at their 99th percentiles. We analyzed the logs of our continuous outcome measures: injury cost, sales, employment, and payroll. To reduce the effect of very small outliers, we added roughly the first percentile of nonzero values to our measures ($79 to Injury cost, 10 to Employment, and $100,000 to Payroll and to Sales) before taking logs; our results were not sensitive to these adjustments (*20*). Summary statistics are reported in table S3.

The preinspection characteristics of treatments and controls were very similar on most measures (e.g., employment, payroll, and sales) (table S4). Whereas the treatments averaged 3.7 injuries per year in the 4-year period preceding the randomized inspection and the controls averaged 3.1 over the same period (*t* test *P* value = 0.06), their pretrends (14% decline for treatments, 12% decline for controls) were statistically indistinguishable (*t* test *P* value = 0.85).

For two reasons, we think that the disparity represents sampling variation rather than conscious selection by Cal/OSHA (*21*). First, we closely replicated Cal/OSHA's random selection procedures to create the pool of establishments at risk of a randomized inspection each year. Second, Cal/OSHA had no information on injury rates for the vast majority of establishments

it randomly inspected. In addition, kernel density plots of several key variables the year before the match year (figs. S1 to S3) revealed nearly identical distributions between the treatments and controls, including the variables for which the statistical tests found significant differences.

Even if due solely to sampling error, this imbalance on preinspection injury rates made it important to adjust for preinspection characteristics in our analysis. Thus, we measured the causal effect of inspections via a difference-in-differences analysis. Specifically, we estimated the following model for each outcome Y_{it} at establishment i in year t:

$$Y_{it} = \alpha_i + \beta \cdot \textit{Has been randomly inspected}_{it} + \Sigma_k \gamma_k \cdot \mathbf{X}_{ikt} + \Sigma_t \delta_t \cdot \textit{year}_t + \varepsilon_{it}$$

where α_i was a complete set of establishment-specific intercepts (or, in some specifications, conditional fixed effects). *Has been randomly inspected*$_{it}$ was coded "1" the year an establishment was randomly inspected and each year thereafter and was otherwise coded "0." Of primary interest is β, which represented the estimated effect of a random inspection; that is, the average change in outcome levels pre- versus postinspection. X_{ikt} referred to controls (subscripted k), such as average occupational riskiness and log employment, that were included in some specifications. All models included a full set of year dummies ($year_t$). The supplementary materials describe multiple robustness checks for each analysis.

We first analyzed the effects of inspections on injury rates and injury cost and then turned to the possible unintended consequences on firm survival, credit ratings, sales, employment, and payroll. To predict the number of injuries at a workplace, we estimated a negative binomial regression model with establishment-level conditional fixed effects. The point estimate in column 1 of Table 1 indicates that randomized inspections reduce annual injuries by 9.4% [β = –0.099, P = 0.013, incident rate ratio = 0.906, 95% CI = –0.177 to –0.021].

The effects of inspections might attenuate after a few years or might take a few years to emerge. To test for such changes in the effects of inspections over time, we replaced the single posttreatment dummy for inspected establishments with a dummy coded "1" only in the randomized inspection year and a series of dummies for each of the subsequent 4 years. Inspections statistically significantly reduced injuries in the random inspection year and 3 and 4 years later, marginally reduced them 1 year later, but had no significant effects 2 years later (column 2). In short, the reduction in injuries after inspections endured. We found nearly identical annual estimates when we excluded matched groups of which either member (the treatment or control) was inspected 3 or 4 years before the match year (table S5).

To extend our analysis beyond average effects, we also estimated distinct effects of these inspections on the number of minor financial claims (resulting in less than $2000 in workers' compensation) and the number of major financial claims (at least $2000). The results of these two regressions were nearly identical: β = –0.107 for smaller claims and β = –0.136 for larger claims, with $P < 0.05$ in both instances (table S6). These results imply that inspections reduce the rates of both minor and major injuries.

Turning to the cost of injuries, an ordinary least squares (OLS) regression model with establishment-level fixed effects indicates that randomly inspected establishments exhibited a 26% decline in injury cost (column 3, β = –0.298, 95% CI = –0.513 to –0.083, exp(β) = 0.74, $P < 0.01$). When we permitted the effect of inspections to differ by years since inspection, the negative point estimates suggested that inspections consistently reduced injury cost, and we could not reject the equality of all these coefficients (P = 0.09, column 4). The pattern of coefficients resembled the pattern for injury rates, with the year-specific treatment effects statistically significant and larger in magnitude in the year of random inspection and years 3 and 4 after the inspection. Results were nearly identical when we excluded matched groups of which either the treatment or control was inspected 3 or 4 years before the match year (table S5).

Table 1. Regressions yield evidence that randomized OSHA inspections reduced workplace injury rate and injury cost (± standard errors). Standard errors clustered by establishment in OLS models (columns 3 and 4). The models in columns 1 and 2 include establishment-level conditional fixed effects. The models in columns 3 and 4 include establishment-level fixed effects. To reduce the effect of very small outliers, we added roughly the first percentile of nonzero values ($79) to Injury cost before taking the log. To reduce the effect of large outliers, Injury count, and Log Injury cost were top-coded at their 99th percentiles. Sample size in columns 1 and 2 is <409 treatments and <409 controls because the negative binomial specification with conditional fixed effects drops establishments that have no variation in their number of injuries.

	(1)	(2)	(3)	(4)
Dependent variable	**Injury count**		**Log Injury cost**	
Specification	**Conditional fixed-effects negative binomial regression**		**Fixed-effects OLS**	
Has been randomly inspected (this year or before)	–0.099 ± 0.040*		–0.298 ± 0.110**	
Year of random inspection		–0.152 ± 0.053**		–0.379 ± 0.123**
One year after random inspection		–0.023 ± 0.055		–0.217 ± 0.145
Two years after random inspection		–0.033 ± 0.063		–0.085 ± 0.172
Three years after random inspection		–0.135 ± 0.077+		–0.558 ± 0.194**
Four years after random inspection		–0.266 ± 0.091**		–0.455 ± 0.223*
Year dummies	Included	Included	Included	Included
Observations (establishment-years)	5593	5593	5872	5872
Number of establishments	765	765	818	818
Number of treatment establishments	389	389	409	409
Number of control establishments	376	376	409	409
Wald tests				
Dependent variable sample mean	3.43	3.43	7.41	7.41
Each treatment coefficient is equal to zero		X^2 = 15.79		F = 3.17
		P = 0.008		P = 0.008
Sum of treatment coefficients equals zero		χ^2 = 7.72		F = 7.13
		P = 0.006		P = 0.008
All treatment coefficients equal to each other		χ^2 = 10.14		F = 2.02
		P = 0.044		P = 0.091

**$P < 0.01$, *$P < 0.05$, +$P < 0.10$.

Table 2. Regressions yielded no evidence that random OSHA inspections influenced employment, payroll, or sales. OLS coefficients ± standard errors clustered by establishment; effects are not statistically significant ($P > 0.10$). To reduce the effect of very small outliers, we added roughly the first percentile of nonzero values (10 to Employment and $100,000 to Payroll and Sales) before taking logs.

	(1)	(2)	(3)
Dependent variable	**Log Employment**	**Log Payroll**	**Log Sales**
Specification	**OLS**	**OLS**	**OLS**
Has been randomly inspected (this year or before)	0.027 ± 0.016	0.005 ± 0.013	0.002 ± 0.044
Year dummies	Included	Included	Included
Establishment-level fixed effects	Included	Included	Included
Observations (establishment-years)	5278	5872	3190
Number of establishments	787	818	640
Number of treatment establishments	390	409	329
Number of control establishments	397	409	311
Dependent variable sample mean	3.61	14.50	14.86

To assess the impact of inspections on workplace survival, we defined an establishment to have "died" if it had disappeared from both the NETS and the WCIRB databases. We were unable to observe if any treatment or control establishments died after the sample period ended in 2006. Fortunately, censoring does not lead to bias because our matching of controls to treatments was in the year of the randomized inspection; thus, data on each matched pair of treatments and controls were right-censored after the identical number of years. In our sample, 4.4% of the treatment establishments did not survive until 2006, a rate slightly but not economically or statistically significantly lower than the 5.6% death rate among the control establishments ($P = 0.423$).

Although treatment status was randomized, there were differences between the treatment and control groups' preinspection sales, employment, and payroll. We ran several specifications of survival analyses that condition on these characteristics using a logit, a conditional logit with a fixed effect for each matched pair, and a Cox proportional hazard model with each matched pair its own strata. For all models, survival rates of randomly inspected establishments were not statistically significantly different from those of the controls (see table S7). These results yielded no support to critics of OSHA who claim that inspections harm companies' survival prospects.

Because company death is relatively rare, we also analyzed whether random inspections affected establishments' creditworthiness, using Dun & Bradstreet's Composite Credit Appraisal and PAYDEX scores. We used ordered logit regression models to predict Composite Credit Appraisal, an ordinal dependent variable that ranged from 1 to 4. We used OLS regression with establishment-level fixed effects to predict minimum PAYDEX scores, which ranged from 1 to 100. The point estimates were positive—hinting that inspections, if anything, increased creditworthiness—but very close to zero and nowhere near statistically significant (table S8).

To assess whether random inspections affect firm growth, we estimated fixed-effect OLS models to predict log employment, log payroll, and log sales (Table 2). Randomly inspected establishments did not differ significantly from controls in employment, total earnings, or sales, although each point estimate was positive. The point estimates show that treatment increases employment and payroll by small amounts (2.7% for employment and 0.5% for payroll, neither statistically significant) with fairly narrow 95% CIs (–0.5% to +5.8% for employment and –2.0% to +3.0% for payroll). Thus, we rule out large declines in employment and payroll. The coefficient on sales was also tiny and positive (0.2%), but the confidence interval was much wider (–8.4% to + 8.8%).

In sum, workplaces that Cal/OSHA randomly inspected (or attempted to randomly inspect) subsequently experienced substantially lower injury rates and workers' compensation costs compared with a matched set of workplaces that were eligible for, but did not receive, a random inspection. The lower injury rates were not transient.

With many assumptions (see supplementary materials), our point estimates imply that the reduction in injuries in the 5 years after a workplace inspection reduced medical costs and lost earnings by roughly $355,000 (in 2011 dollars) (*22*–*24*). This estimated 5-year total is ~14% of the average annual payroll of this sample of employers. Thus, although admittedly imprecise, the estimated benefits of a randomized safety inspection appear to be substantial. These results do not support the hypothesis that OSHA regulations and inspections on average have little value in improving health and safety.

Although this estimated value of improved health is fairly large, it is crucial to know how much employers pay for these improvements in safety, as well as how much employees pay in terms of lower wages or employment. As noted above (and formalized in an illustrative model in the supplementary materials), economists' benchmark model suggests that the increased costs of safety measures that reduce injury rates can also reduce wages, employment, and rates of firm survival. Although we cannot rule out any of these unintended consequences, we found no evidence that inspections lead to worse outcomes for employees or employers. The point estimates on changes in employment, payroll, sales, and credit ratings were all positive, although all coefficients were small, and none approached statistical significance.

The estimates in Table 2 imply that we can be 95% certain that the mean establishment either grows payroll or experiences a decline of less than $221,000 over the 5 years after the inspection (*25*). The lower-bound estimate of lost payroll is in different units than lost earnings and medical costs, and there is substantial uncertainty about our estimated benefits (with a point estimate of $355,000). With that said, these calculations imply that employees almost surely gain from Cal/OSHA inspections.

This result is not consistent with the perfectly competitive model's prediction that Cal/OSHA's mandated increases in safety would reduce employment and/or earnings sufficiently that, on average, employees would be worse off. These results therefore suggest that it is important to test which assumptions of the perfectly competitive model are sufficiently violated to drive this result (e.g., that employees have very good information on hazards or that labor is perfectly mobile).

Our study has several limitations, including its focus on a subset of companies (single-establishment firms in high-hazard industries and with at least 10 employees) in one region (California), a single type of inspection (randomized, not those driven by complaints or by serious accidents), and a single workplace-safety regulator (Cal/OSHA). Our method also ignores any effects of the threat of inspections on as-yet-uninspected workplaces. It is important to replicate this study in other settings and by using additional study designs to examine the generalizability of our results. It is also important to supplement statistical studies such as this one with qualitative research that helps us understand the process by which workplace regulations affect (and sometimes improve) outcomes.

References and Notes

1. A. Haviland, R. Burns, W. Gray, T. Ruder, J. Mendeloff, *J. Safety Res.* **41**, 339 (2010).
2. J. Feldman, *OSHA Inaction: Onerous Requirements Imposed on OSHA Prevent the Agency from Issuing Lifesaving Rules* (Public Citizen's Congress Watch, Washington, DC, 2011): www.citizen.org/documents/osha-inaction.pdf.
3. S. Pelley, Is enough done to stop explosive dust? *60 Minutes* (CBSnews.com), June 8, 2008, www.cbsnews.com/stories/2008/06/05/60minutes/main4157170.shtml.
4. MSDSonline, New safety poll: Do OSHA regulations kill jobs or stop jobs from killing people? October 13, 2011, http://blog.msdsonline.com/2011/10/new-safety-poll-do-osha-regulations-kill-jobs-or-stop-jobs-from-killing-people/.
5. J. Sherk, Opportunity, parity, choice: A labor agenda for the 112th Congress (Heritage Foundation, Washington DC, July 14, 2011), www.heritage.org/research/reports/2011/07/opportunity-parity-choice-a-labor-agenda-for-the-112th-congress.

6. Public Citizen's Congress Watch, Sen. Coburn is dead wrong on worker safety [press release] (1 August 2011), www.citizen.org/pressroom/pressroomredirect.cfm?ID = 3394.
7. R. S. Smith, *Ind. Labor Relat. Rev.* **32**, 339 (1979).
8. The supplementary materials provide a formal model of these results.
9. R. S. Smith, *J. Hum. Resour.* **14**, 145 (1979).
10. W. K. Viscusi, *Bell J. Econ.* **10**, 117 (1979).
11. J. W. Ruser, R. S. Smith, *J. Hum. Resour.* **26**, 212 (1991).
12. W. B. Gray, J. T. Scholz, *Law Soc. Rev.* **27**, 177 (1993).
13. J. Mendeloff, W. Gray, *Law Policy* **27**, 219 (2005).
14. A. Haviland, R. M. Burns, W. B. Gray, T. Ruder, J. Mendeloff, *Am. J. Ind. Med.* 10.1002/ajim.22062 (2012).
15. W. B. Gray, *Am. Econ. Rev.* **77**, 998 (1987).
16. C. Dufour, P. Lanoie, M. Patry, *J. Prod. Anal.* **9**, 233 (1998).
17. J. W. Ruser, *J. Risk Insur.* **62**, 67 (1995).
18. J. P. Leigh, J. P. Marcin, T. R. Miller, *J. Occup. Environ. Med.* **46**, 10 (2004).
19. California Department of Industrial Relations, *2005 Report on the High Hazard Enforcement Program and High Hazard Consultation Program* (Division of Occupational Safety and Health, Sacramento, CA, 2007), www.dir.ca.gov/dosh/enforcementpage.htm, accessed September 2011.
20. To assess the impact of our handling of outliers, we reestimated our models on variables that were neither top-coded to correct for large outliers nor corrected to account for very small outliers (i.e., we added only 1 before taking the log of Injury count, Injury cost, Sales, Employment, and Payroll). The results, presented in tables S9 and S10 in the supplementary materials, continue to indicate that inspections lead to statistically significant reductions in Injury count and Injury cost. The results also continue to yield no evidence that inspections affected Employment, Payroll, or Sales. The magnitude of these estimated effects on injury rates and injury costs exceeded those yielded by our primary model results. This confirms the conservative nature of our primary estimates and suggests the importance of mitigating the influence of outliers.
21. Cal/OSHA would have had some data on injury rates for workplaces they had recently inspected. Because their procedures were to avoid randomized inspections for workplaces with any inspection in the previous 2 years, we dropped potential treatments and controls that had been inspected within 2 years before the match year. Cal/OSHA only had inspected 7% of treatments in the 4 years before the random inspection year; results were unchanged when we dropped treatments or controls with inspections in the prior 4 years.
22. In the supplementary materials, we show that the cost of reported injuries in medical care and lost wages, not counting pain and suffering, is very roughly $8400 per employee in high-hazard industries in California. If an inspection reduces all costs by the same 26% that we estimated for workers' compensation costs (Table 1, column 3) and if there is an average of 33 employees per employer in our sample, then a Cal/OSHA inspection leads to roughly $71,000 in lower medical costs and lost wages per year. If the effect lasts from the inspection year through the next 4 years (as in Table 1, column 4), the total value to society of an inspection is very approximately on the order of $355,000. This estimate is very rough and ignores the underreporting of injuries (*23*, *24*), safety benefits lasting more than 4 years, the reduction in pain and suffering, and (working in the opposite direction) the discounting of future benefits.
23. K. D. Rosenman *et al.*, *J. Occup. Environ. Med.* **42**, 25 (2000).
24. J. Biddle, K. Roberts, K. D. Rosenman, E. M. Welch, *J. Occup. Environ. Med.* **40**, 325 (1998).
25. The figure of $221,000 is the lower bound of the 95% CI of our estimate on payroll (–0.021) times mean payroll ($2,101,500) times 5 years.

Acknowledgments: We appreciate the support of C. Baker and I. Nemirovsky of California's Commission on Health and Safety and Workers' Compensation, D. Bellusci and T. Basuino of the Workers' Compensation Insurance Rating Board, and G. Murphy of California's Division of Occupational Safety and Health and helpful comments from F. Neuhauser and J. Sum; A. Nguyen-Chyung and M. Ouellet provided outstanding research assistance. We are grateful for financial support from the Commission on Health and Safety and Workers' Compensation, Harvard Business School's Division of Research and Faculty Development, Kauffman Foundation, University of California at Berkeley's Institute for Research on Labor and Employment, and University of California's Labor and Employment Relations Fund. The publicly accessible portion of the data set analyzed in this study is available at http://hdl.handle.net/1902.1/17936 in the Murray Research Archive at Harvard University (study identifier: 1902.1/17936). That data set contains the following variables: a scrambled establishment ID that uniquely identifies each establishment, the establishment's city, industry, year, year of random inspection, treated (has been randomly inspected), sales, employment, PAYDEX score, and Composite Credit Appraisal. It does not contain the following variables used in the analysis because of the confidentiality conditions under which they were obtained: establishment name, street address, ZIP Code, DUNS number; annual payroll, injury count, injury cost, and average occupational riskiness. Details on how to access all data used in this analysis are provided in the supplementary materials.

Supplementary Materials

www.sciencemag.org/cgi/content/full/336/6083/907/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S3
Tables S1 to S12
References (*26*–*33*)

12 October 2011; accepted 4 April 2012
10.1126/science.1215191

Cost-Benefit Tradeoffs in Engineered *lac* Operons

Matt Eames[1] and Tanja Kortemme[1,2]*

Cells must balance the cost and benefit of protein expression to optimize organismal fitness. The *lac* operon of the bacterium *Escherichia coli* has been a model for quantifying the physiological impact of costly protein production and for elucidating the resulting regulatory mechanisms. We report quantitative fitness measurements in 27 redesigned operons that suggested that protein production is not the primary origin of fitness costs. Instead, we discovered that the *lac* permease activity, which relates linearly to cost, is the major physiological burden to the cell. These findings explain control points in the *lac* operon that minimize the cost of *lac* permease activity, not protein expression. Characterizing similar relationships in other systems will be important to map the impact of cost/benefit tradeoffs on cell physiology and regulation.

Expressing proteins uses cellular resources and thus incurs fitness costs (*1*, *2*). To balance these costs and generate a net fitness advantage, cells must couple protein expression to beneficial processes. These cost/benefit tradeoffs (*3*) shape mechanisms that regulate protein expression, such as those in the *lac* operon (*4*). The fitness costs of protein expression have also been hypothesized to govern the speed at which proteins evolve (*5*, *6*) and to influence the operation of regulatory circuits (*7*, *8*). To interpret these effects and derive predictive models of the physiological consequences of protein expression, the underlying sources of both cost and benefit must be identified and quantified. Such models are central to understanding gene regulation, metabolic engineering, and molecular evolution.

Because costs are balanced or even completely masked by coupled benefits under physiological conditions, cost and benefit can be difficult to separate. We used the *lac* operon (*4*, *9*) (Fig. 1A) to separately quantify the cost and benefit of protein expression (*3*); we define cost as the relative reduction in growth rate due to operon expression and benefit as the relative increase in growth rate in the presence of lactose, the substrate of the operon. To dissect the interplay between proposed cost sources and protein benefit, we quantified the effects of genetic changes that modulate three cost/benefit tradeoffs (*5*): protein production efficiency (*10*) (by changing translational optimization and thereby expression level), functional efficiency (by modulating catalysis), and folding efficiency (*6*) (by altering the propensity to misfold).

To determine the growth response, we induced expression of the *lac* operon using the nonmetabolized inducer isopropyl-β-D-thiogalactopyranoside (IPTG) and varied the concentration of lactose. At low lactose concentrations, the change in growth rate relative to that of uninduced cells is assumed to primarily reflect the cost of protein expression, whereas at higher lactose concentrations, growth reflects both the cost and benefit of lactose metabolism. We performed our experiments at full induction to decouple regulatory effects from the cost and benefit of expression and also to avoid complications arising from bistability at low inducer concentrations (*1*). We knocked out the entire *lac* operon and replaced it with engineered versions at the attTn7 locus (*11*). As a control, we confirmed that a knockin (*KIlac*) of the wild-type *lac* operon successfully recaptured native cost/benefit lactose response curves (fig. S1).

[1]Graduate Group in Biophysics, MC 2530, University of California, San Francisco, CA 94158–2330, USA. [2]California Institute for Quantitative Biomedical Research and Department of Bioengineering and Therapeutic Sciences, MC 2530, University of California, San Francisco, CA 94158–2330, USA.

*To whom correspondence should be addressed. E-mail: kortemme@cgl.ucsf.edu

Previous work attributed the cost of expressing the *lac* operon to the process of producing its proteins, with the majority of the cost thought to result from *lacZ* expression (*2*, *3*) (Fig. 1A). To quantify the effect of varying protein production cost by genetic changes, we constructed eight different strains with altered codon usage in *lacZ* (Table 1) and measured their lactose response curves (Fig. 1B). These strains reduced the activity of LacZ to levels between 6 and 51% of that of *KIlac* (Table 1) and spanned a range of cost and benefit (Fig. 1B).

The inability of the redesigned production efficiency strains to express at levels greater than that of *KIlac* was unexpected, but consistent with the complex relationship between nucleotide sequence and expression (*10*, *12*). However, independent of this complexity, the range of expression levels allowed us to directly relate protein activity to cost. Plotting LacZ levels against growth cost for the production efficiency strains revealed that after an initial rise, the cost of expression leveled off as LacZ activity (as a proxy for expression) increased (Fig. 2A). This was surprising because, if LacZ expression was the primary cost, we would instead predict that costs would increase further with increasing expression.

To measure how the functional efficiency of LacZ affects cost and benefit in lactose response curves, we tested five engineered strains with point mutations that represent a range of LacZ catalytic activities (Table 1). This revealed a gradual loss of LacZ benefit in these strains (Fig. 1C) in order of decreasing activity, consistent with the assumed relationship between benefit and LacZ activity. However, unexpectedly, in the *E537Q* strain encoding a severely catalytically comprised LacZ mutant, the cost increased as a function of lactose concentration (Fig. 1C). We verified that the E537Q mutation did not affect expression levels by determining LacZ protein levels of green fluorescent protein–tagged strains of *KIlac* and *E537Q* (fig. S2). The increasing cost in the *E537Q* strain suggested a penalty previously masked in the lactose response curves.

Because the mutations altering functional efficiency were engineered to minimize changes in protein stability, the observed penalty in the *E357Q* strain reflects activity-dependent effects. To determine possible additional effects of folding efficiency, we engineered two mutations with reduced stability (though also observed to have reduced catalytic performance) and three mutations predicted by structure-based modeling to have destabilization effects (*11*). The lactose response curves (Fig. 1D) of four of these strains exhibited the same lactose-dependent penalty as the *E537Q* strain. We also did not observe substantial cost differences at minimal lactose concentrations between mutations affecting catalytic performance (Fig. 1C) and mutations affecting both stability and catalytic performance (Fig. 1D). Furthermore, the addition of a tag that targets LacZ for degradation by the ClpXP and ClpAP proteases (*11*) and reduces intracellular LacZ levels by a third (the *degtag* strain, Table 1) did not alter the cost or benefit of protein production measurably (Fig. 1D). These results support the idea that potential costs of destabilized proteins and increased activation of degradation pathways are not detectable in our system (*2*).

The observed benefit in the engineered strains follows an expected relationship with LacZ activity. Plotting the difference in relative growth rate between maximum and minimum lactose concentrations (benefit) against measured LacZ activity (Fig. 2B), we see that the benefit increases proportionally to LacZ activity until levels reach approximately 30% of that of *KIlac*; after this point no additional benefit is conferred. This value is consistent with observations that cells grown on lactose exhibit only 20 to 40% of the LacZ activity found in fully induced cells (*13*). In contrast, the effects of protein production, functional, and folding efficiency on cost revealed that at high LacZ levels, protein expression costs do not increase as much as expected even though more LacZ is produced (Fig. 2A). This cannot be explained by models that attribute the majority of the cost of expressing the *lac* operon to the production of LacZ, the largest protein in the operon. These models predicted

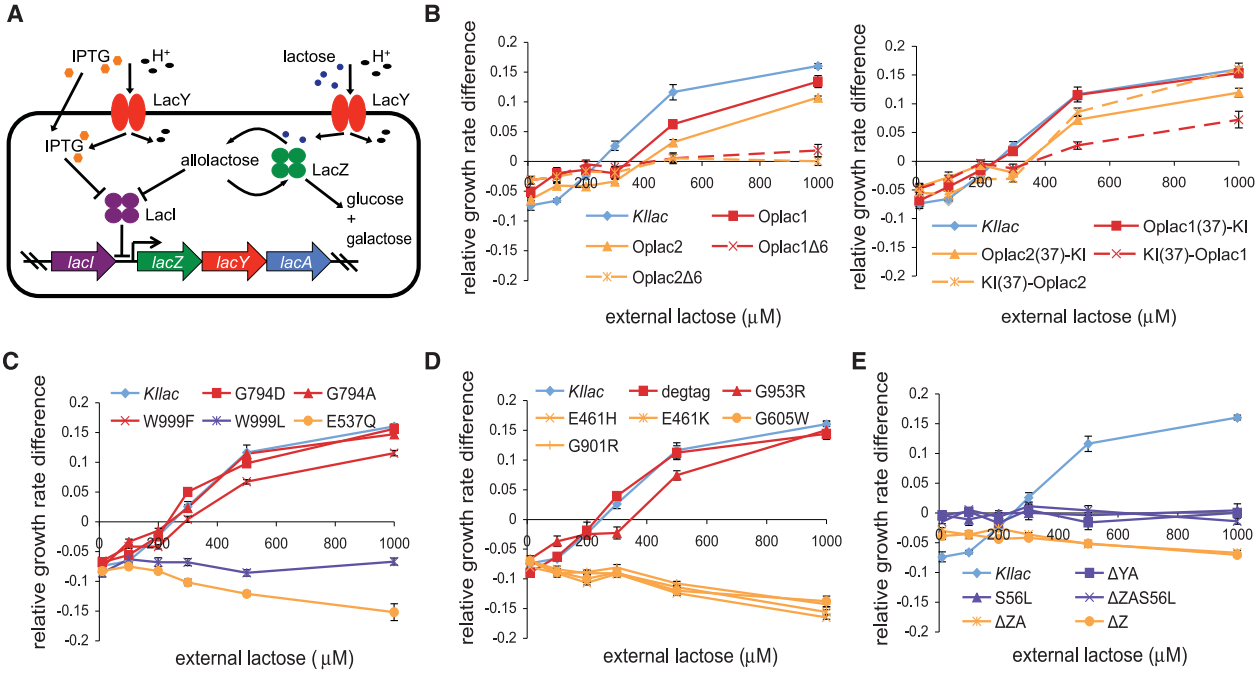


Fig. 1. Cost/benefit measurements in the *lac* operon. (**A**) *lac* operon model system. LacI: *lac* repressor inhibiting operon expression. LacZ: β-galactosidase, which converts lactose either into glucose and galactose to generate metabolic benefit or into allolactose, which induces operon expression by relieving LacI repression; the artificial inducer IPTG can also induce operon expression. LacY: lactose/proton symporter; LacA: transacetylase. (**B** to **E**) Cost/benefit tradeoffs in engineered strains, quantified as the growth rate difference of IPTG-induced relative to uninduced populations at varying lactose (lactose response curves). Cost was initially assumed to result largely from LacZ expression and to dominate at low lactose concentrations, but to be balanced by benefit with increasing amounts of lactose. (B) Lactose response curves of strains (codon-redesigned *lacZ*, Table 1) with varying production efficiency. (C) Lactose response curves for strains (point mutations affecting LacZ catalytic parameters, Table 1) with varying functional efficiency. (D) Lactose response curves for strains (point mutations as indicated are designed to destabilize LacZ) with varying folding efficiency. (E) Lactose response curves of strains with minimized LacZ or LacY activity (Table 1). In each panel, the lactose response curve of *KIlac* is shown for reference. For all graphs, bars represent the standard error (SE) of 172 measurements.

that costs should increase nonlinearly (not level off) as more cellular resources are redirected toward higher expression of LacZ. We also observed an unexplained lactose-dependent penalty in strains with severely compromised LacZ activity (Fig. 1, C and D).

To investigate why our data were inconsistent with previous models, we first determined the effect of LacZ expression. We engineered and tested a Δ*YA* strain (knocking out *lacY* and *lacA*) and found that it produced levels of LacZ essentially the same as those of *KIlac* at a minimal cost to the cell (Fig. 1E and Table 1). Therefore, costs associated with LacZ production are not the dominant source of cost. The lack of a growth disadvantage in the Δ*YA* strain suggested that there is a cost unrelated to LacZ expression. To test for relationships between the membrane permease LacY and cost, suggested in earlier work that explored the selective pressure on LacY (*14*), we determined levels of LacY activity by measuring the import of radiolabeled lactose (*11*) by strains in Table 1. These measurements revealed a linear relationship between LacY activity and the relative reduction in growth rate (Fig. 2C). It appears that the modulation of codon usage in *lacZ* in the redesigned production efficiency strains profoundly influences *lacY* expression. This expression polarity, i.e., that attenuation of LacZ levels reduces the expression of downstream genes, has been reported previously (*15*) (fig. S3).

To determine whether the observed cost was due to the production/membrane insertion of LacY (*16*, *17*) or more specifically to its function, we introduced a mutation (S56L) to LacY known to disrupt permease functionality but not affect LacY levels (*18*). We observed that growth cost in the *S56L* strain was essentially eliminated, whereas LacZ activity was unaffected (Fig. 1E, Table 1). This suggests that the activity, not production, of LacY plays the dominant role in the cost of *lac* operon expression.

To test whether the observed lactose-dependent penalty (Fig. 1, C and D) is also dependent on LacY activity, we generated a Δ*ZA* strain (Table 1). Measuring the lactose response curves of Δ*ZA* and Δ*ZA-S56L*, we observed that only Δ*ZA* maintained the penalty (Fig. 1E; the penalty is reduced because of decreased LacY levels in the Δ*ZA* strain). Taken together, these results support the idea that the lactose-dependent penalty is linked only with functional LacY and disappears when LacY is still expressed but not functional.

Both the observed cost and the lactose-dependent penalty (Fig. 1, C and D) could be explained by a reduction in the proton motive force when lactose is transported by the proton/

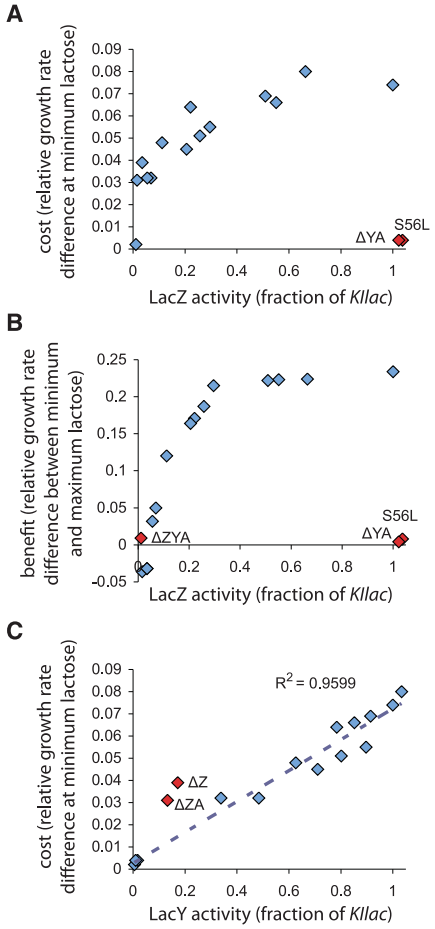


Fig. 2. Dependence of cost and benefit on protein activity for strains in Table 1. (**A**) Plot of cost against LacZ activity. Strains with red symbols and labels have minimal LacY activity. (**B**) Plot of benefit against LacZ activity. (**C**) Plot of cost against LacY activity. Strains with red symbols, Δ*ZA* and Δ*Z*, have minimal LacZ activity, which may result in an artificially low measurement of LacY activity as more ^{14}C-lactose is exported instead of metabolized. Fitting a linear relationship to the data (not including Δ*ZA* and Δ*Z*) yields the equation: cost = 0.0697(LacY activity) + 0.0026, R^2 = 0.9599 (*R*, Pearson correlation coefficient).

Table 1. Designed strains (*11*) with measured LacZ activity and LacY activity (with their standard errors), or catalytic parameters. k_{cat}, turnover number; K_m, Michaelis constant. The folding efficiency strains are listed in Fig. 1D.

Production efficiency strains

Strain	Design (synonymous changes in *lacZ*)	LacZ activity (fraction of that of *KIlac*)	LacY activity (fraction of that of *KIlac*)
KIlac	Knockin reference strain	1	1
ΔZYA	Knockout	0.01 ± 0.01	0.01 ± 0.00
WT	Wild-type operon	0.55 ± 0.07*	0.85 ± 0.03*
Oplac1	Codon-optimized *lacZ*†	0.26 ± 0.02	0.80 ± 0.01
Oplac2	Codon-optimized *lacZ*‡	0.22 ± 0.07	0.78 ± 0.00
Oplac1Δ6	*Oplac1* starting at base 7	0.07 ± 0.02	0.48 ± 0.03
Oplac2Δ6	*Oplac2* starting at base 7	0.06 ± 0.03	0.34 ± 0.03
Oplac1 (37)-KI	*KIlac* with first 37 bases from *Oplac1*§	0.51 ± 0.07	0.91 ± 0.03
Oplac2 (37)-KI	*KIlac* with first 37 bases from *Oplac2*§	0.21 ± 0.4	0.71 ± 0.03
KI (37)-Oplac1	*Oplac1* with first 37 bases from *KIlac*§	0.11 ± 0.01	0.63 ± 0.06
KI (37)-Oplac2	*Oplac2* with first 37 bases from *KIlac*§	0.30 ± 0.02	0.90 ± 0.05

Strains to dissect origin of the cost

Strain	Design	LacZ activity (fraction of that of *KIlac*)	LacY activity (fraction of that of *KIlac*)
degtag	*lacZ* C-terminal *sssA* degradation tag	0.66 ± 0.10	1.03 ± 0.02
ΔZ	*lacZ* deletion	0.03 ± 0.02	0.17 ± 0.00
ΔZA	*lacZ* and *lacA* deletion	0.02 ± 0.01	0.13 ± 0.01
ΔYA	*lacY* and *lacA* deletion	1.02 ± 0.22	0.01 ± 0.00
S56L	Impaired LacY function	1.04 ± 0.07	0.02 ± 0.01

Functional efficiency strains‖

Strain	k_{cat} (s^{-1})	K_m (mM)
WT	600	0.12
G794D	285	0.16
G794A	189	0.19
W999F	54	0.26
W999L	20	3.3
E537Q	0.003	0.12

*LacZ and LacY activity levels of the wild-type operon are reduced relative to that of *KIlac*, most likely because *KIlac* is integrated at a different chromosomal locus (*11*). All other strains used in this work are integrated at the same locus as *KIlac* to allow direct comparison. †Sequence with preferred codons at every position (table S1). ‡Sequence that incorporates codons for a subset of amino acids read by transfer RNAs that are the most highly charged during amino acid starvation (table S1). §Recombination of sequences in regions near the ribosome binding site (*10*, *11*). ‖Catalytic parameters were taken from the literature (*11*).

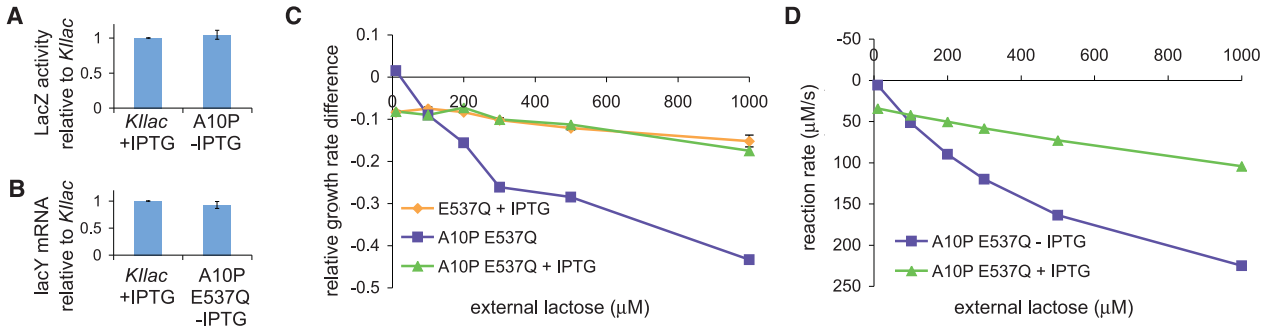


Fig. 3. Dependence of cost on the transport of lactose and IPTG. (**A**) Measured LacZ activity in *KIlac* (induced with IPTG) and the constitutive *A10P* strain (in the absence of IPTG) (*11*). (**B**) Measured *lacY* mRNA levels in *KIlac* (induced with IPTG) and the constitutive *A10P-E537Q* strain (in the absence of IPTG) (*11*). Error bars in (A) and (B) are the standard deviation of three experiments. (**C**) *A10P-E537Q* lactose response curves in the absence or with the addition of IPTG. (**D**) Dependence of reaction rate v of transport through LacY on lactose concentration, predicted from a simple model of competing substrates: $v = (V_L[\text{lactose}]/K_{mL}) + (V_I[\text{IPTG}]/K_{mI})/(1 + [\text{lactose}]/K_{mL} + [\text{IPTG}]/K_{mI})$, where V_L and V_I are the maximum velocity (V_{max}) for lactose and IPTG, respectively, and K_{mL} and K_{mI} are the Michaelis constant (K_m) for lactose and IPTG, respectively. We estimated K_{mL} and V_L from ranges in the literature to be 1 mM and 0.6 mM/s, respectively, and estimated K_{mI} and V_I to be 0.25 mM and 0.1 mM/s, respectively (*11*), assuming reduced efficiency for the non-native substrate (the model recaptured the IPTG-based inhibition of proton symport for a range of estimated IPTG parameters).

lactose symporter LacY (*19*). However, we also detected an expression cost in all strains containing active LacY even at very low lactose concentrations (Fig. 1, B to D). To test the idea that IPTG/proton symport and not permease leakiness was the source of this cost in minimal lactose, we measured growth rate differences at varying IPTG concentrations (fig. S4). We observed increased expression costs with increased IPTG concentration, despite minimal changes in expression levels (the IPTG concentration is high enough for full induction). This suggested that IPTG/proton symport contributed to the observed expression costs.

To determine the dependence of cost on lactose concentration in the absence of both IPTG and the benefit associated with lactose metabolism, we built a strain with a point mutation (*A10P*) in the *lac* repressor to constitutively express the *lac* operon in the absence of IPTG. LacZ activity levels confirmed that *A10P* without IPTG expressed the *lac* operon at levels comparable to those of IPTG-induced *KIlac* (Fig. 3A). We then engineered the A10P mutation into a strain containing the LacZ-inactivating E537Q mutation and confirmed that this strain has a similar expression level to *KIlac* by measuring its *lacY* mRNA level (Fig. 3B). As expected from a model where cost depends on transport through LacY, full expression of the *lac* operon only incurred minimal costs under minimal lactose and no IPTG (Fig. 3C), whereas cost increased with increasing lactose. However, under higher lactose concentrations, the cost was larger in the *A10P-E537Q* strain (without IPTG) than in the *E537Q* strain (with IPTG). This suggested that IPTG transport through the *lac* permease, although costly to the cell, is less efficient than and may competitively inhibit lactose transport. To test whether IPTG indeed slows lactose transport and hence decreases cost, we added IPTG to the *A10P-E537Q* strain. As predicted, this resulted in a lactose response curve indistinguishable from that of *E537Q*. We also confirmed that the observed relationship between IPTG and lactose transport and cost could be recapitulated with a simple model of competitive substrates (Fig. 3D).

Our approach to perturb production, functional, and folding efficiency in redesigned strains to quantify cost and benefit led to the discovery that LacY activity is a major physiological source of expression costs in the *lac* operon. Although this result was unexpected, on the basis of previous interpretations of the cost/benefit relationships in the *lac* model system (*2*, *3*) and common assumptions about the cost of protein expression (*3*), our findings are consistent with observations that cells taken from a lactose-limited chemostat die when plated on lactose minimal medium; previously denoted as "lactose killing" (*19*). The quantitative relationship we derive between LacY activity and fitness (Figs. 2C and 3D) shows that the LacY activity–dependent effects outweigh those of protein production and misfolding in the *lac* operon. Therefore, although production and misfolding costs (*6*, *20*, *21*) may of course generally apply to many proteins (supplementary materials text), these costs are small compared to the activity costs of the LacY membrane permease that dominate the cost/benefit tradeoffs in the *lac* operon model.

Our findings also reinterpret the significance of regulation mechanisms (in addition to the LacI-based repression of operon expression, Fig. 1A) to minimize the physiological costs of LacY activity and couple them to accompanying benefits. Protein activity–associated costs predict that control directly at the protein (and not the gene) level is advantageous. This regulation is realized by the direct inactivation of LacY function in the presence of glucose and other carbon sources, known as "inducer exclusion" (*22*). The use of allolactose (produced by a secondary activity of LacZ, Fig. 1A) as the natural inducer of the operon (*23*), and not lactose, provides the advantage of verifying beneficial LacZ functionality before expressing the operon and its costly LacY component.

Some of these layers of regulation and compensatory effects, as well as the dependence of fitness on the energy state of the cell, may at least in part explain why the LacY-dependent activity costs have previously been missed (supplementary materials text). The means by which we generated our reduced-expression strains may have also enabled us to observe these effects: We altered the codon usage of *lacZ* whereas prior studies have controlled LacZ expression levels with IPTG. These differences will not only generate costs through proton symport but will also differentially affect downstream expression of *lacY*. In addition, direct quantification of the differential LacY activity in our engineered strains revealed its linear relationship with cost.

The cost of expressing functional LacY suggests that there may be selective pressure to group *lac* genes into an operon, so that LacZ benefit offsets LacY costs through strict coexpression. Future experimentation could attempt to tease apart the cost/benefit contributions of functional proteins co-occurring in the same pathway to elucidate the source of similar fitness tradeoffs as uncovered here in the *lac* operon model.

References and Notes

1. A. Novick, M. Weiner, *Proc. Natl. Acad. Sci. U.S.A.* **43**, 553 (1957).
2. D. M. Stoebel, A. M. Dean, D. E. Dykhuizen, *Genetics* **178**, 1653 (2008).
3. E. Dekel, U. Alon, *Nature* **436**, 588 (2005).
4. F. Jacob, J. Monod, *J. Mol. Biol.* **3**, 318 (1961).
5. D. A. Drummond, J. D. Bloom, C. Adami, C. O. Wilke, F. H. Arnold, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 14338 (2005).
6. D. A. Drummond, C. O. Wilke, *Cell* **134**, 341 (2008).
7. S. Klumpp, Z. Zhang, T. Hwa, *Cell* **139**, 1366 (2009).
8. M. Scott, C. W. Gunderson, E. M. Mateescu, Z. Zhang, T. Hwa, *Science* **330**, 1099 (2010).
9. J. R. Beckwith, in Escherichia coli *and* Salmonella typhimurium*: Cellular and Molecular Biology*,

F. C. Neidhardt, Ed. (American Society for Microbiology, Washington, DC, 1987), pp. 1444–1452.
10. G. Kudla, A. W. Murray, D. Tollervey, J. B. Plotkin, *Science* **324**, 255 (2009).
11. Materials and methods are available as supplementary materials on *Science* Online.
12. A. Hui, J. Hayflick, K. Dinkelspiel, H. A. de Boer, *EMBO J.* **3**, 623 (1984).
13. D. M. Wilson, M. Kusch, J. L. Flagg-Newton, T. H. Wilson, *FEBS Lett.* **117** (suppl.), K37 (1980).
14. D. E. Dykhuizen, A. M. Dean, D. L. Hartl, *Genetics* **115**, 25 (1987).
15. O. Yarchuk, N. Jacques, J. Guillerez, M. Dreyfus, *J. Mol. Biol.* **226**, 581 (1992).
16. D. Dykhuizen, M. Davies, *Ecology* **61**, 1213 (1980).
17. A. L. Koch, *J. Mol. Evol.* **19**, 455 (1983).
18. A. L. Green, E. J. Anderson, R. J. Brooker, *J. Biol. Chem.* **275**, 23240 (2000).
19. D. Dykhuizen, D. Hartl, *J. Bacteriol.* **135**, 876 (1978).
20. D. A. Drummond, C. O. Wilke, *Nat. Rev. Genet.* **10**, 715 (2009).
21. K. A. Geiler-Samerotte *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 680 (2011).
22. T. Osumi, M. H. Saier Jr., *Proc. Natl. Acad. Sci. U.S.A.* **79**, 1457 (1982).
23. A. Jobe, S. Bourgeois, *J. Mol. Biol.* **69**, 397 (1972).

Acknowledgments: We thank C. Gross, A. Sali, B. Simmons, J. Tabor, R. Oberdorf, R. Ritterson, D. Stoebel, D. Dykhuizen, U. Alon, E. Dekel, A. Drummond, K. Geiler, J. Fraser, and A. Stein for discussions and comments, and McKenzie and Craig for the pGRG37 vector. This work was supported by the Sandler Program for Breakthrough Biomedical Research. Data are available in the supplementary materials.

Supplementary Materials
www.sciencemag.org/cgi/content/full/336/6083/911/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S7
Tables S1 to S3
References (*24*–*41*)

13 January 2012; accepted 3 April 2012
10.1126/science.1219083

How Hibernation Factors RMF, HPF, and YfiA Turn Off Protein Synthesis

Yury S. Polikanov,[1,3]* Gregor M. Blaha,[1]* Thomas A. Steitz[1,2,3]†

Eubacteria inactivate their ribosomes as 100*S* dimers or 70*S* monomers upon entry into stationary phase. In *Escherichia coli*, 100*S* dimer formation is mediated by ribosome modulation factor (RMF) and hibernation promoting factor (HPF), or alternatively, the YfiA protein inactivates ribosomes as 70*S* monomers. Here, we present high-resolution crystal structures of the *Thermus thermophilus* 70*S* ribosome in complex with each of these stationary-phase factors. The binding site of RMF overlaps with that of the messenger RNA (mRNA) Shine-Dalgarno sequence, which prevents the interaction between the mRNA and the 16*S* ribosomal RNA. The nearly identical binding sites of HPF and YfiA overlap with those of the mRNA, transfer RNA, and initiation factors, which prevents translation initiation. The binding of RMF and HPF, but not YfiA, to the ribosome induces a conformational change of the 30*S* head domain that promotes 100*S* dimer formation.

Bacteria slow down protein synthesis during nutrient starvation by converting ribosomes into translationally inactive 100*S* dimers (*1*) or 70*S* monomers (*2*), enter a stationary phase, and cease to grow. In this phase, bacterial cells are resistant to external stresses, which allows them to resist antimicrobial agents (*3*) and to engage in increased mutagenesis (*4*). Expression of stationary-phase proteins, such as ribosome modulation factor (RMF) (*1*) and hibernation promoting factor (HPF) (*5*), results in formation of 100*S* dimer, which leads to "ribosome hibernation" that aids cell survival.

In *Escherichia coli* (*Eco*) RMF and HPF proteins are encoded by the *rmf* and *hpf* genes, respectively. In contrast to the *rmf* gene, deletion of the *hpf* gene neither affects cell viability nor affects dimer formation in vivo during the stationary phase (*5*). The binding of RMF both in vivo and in vitro causes dimerization of 70*S* ribosomes into 90*S* particles (*6*), which are then stabilized as 100*S* dimers upon HPF binding (*6*). The 90*S* and 100*S* particles are both dimers but have different sedimentation coefficients. Negative staining (*7*) and cryo–electron microscopy (cryo-EM) studies (*8*, *9*) showed that the two 70*S* ribosomes in a 100*S* dimer are linked together via their small subunits in vivo. Unlike RMF, HPF alone cannot induce ribosome dimerization (*6*).

An additional stationary-phase protein, YfiA, encoded by the *yfiA* gene, promotes the formation of translationally inactive monomeric 70*S* ribosomes (*10*), which consequently prevents the recycling of ribosomes for translation initiation. Although YfiA and HPF share a 40% sequence similarity (*6*), they have different effects on 100*S* dimer formation: HPF converts 90*S* into 100*S* particles, whereas YfiA prevents RMF-dependent 90*S* formation (*6*). The low-resolution electron density maps of YfiA (11.5 Å) and PSRP1 (8.5 Å), a chloroplast-specific YfiA homolog, bound to the ribosome, suggest that they can interfere with protein synthesis because their binding sites overlap with those of transfer RNAs (tRNAs) on the 30*S* subunit (*11*, *12*). Although the structural models of these stationary-phase proteins exist, high-resolution structures of each of these proteins in complex with the ribosome are required to understand their mechanisms of action.

Here, we present three high-resolution crystal structures of the *Thermus thermophilus* (*Tth*) 70*S* ribosome in complex with RMF, HPF, or YfiA that were refined by using data extending to 3.0 Å ($I/\sigma I = 1$), 3.1 Å ($I/\sigma I = 1$), and 2.75 Å ($I/\sigma I = 1$) resolution, respectively. The resolutions at which

[1]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520–8114, USA. [2]Department of Chemistry, Yale University, New Haven, CT 06520–8114, USA. [3]Howard Hughes Medical Institute, Yale University, New Haven, CT 06520–8114, USA.
*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: thomas.steitz@yale.edu

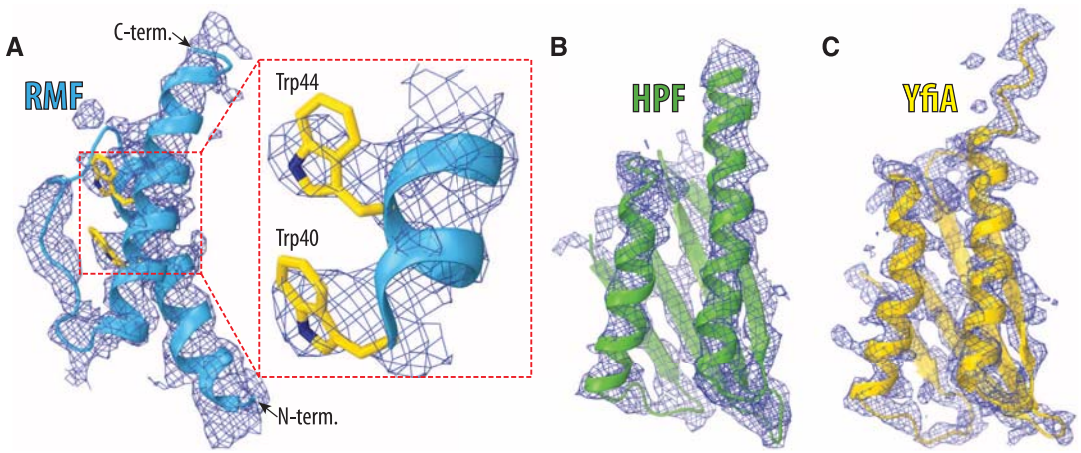


Fig. 1. Unbiased ($F_{obs} - F_{calc}$) difference Fourier maps of hibernation factors in complex with the *T. thermophilus* 70*S* ribosomes. The unbiased difference electron densities are contoured at 2.2, 2.5, and 2.7 σ for RMF (**A**), HPF (**B**) and YfiA (**C**), respectively. The refined models of RMF (blue), HPF (green), and YfiA (yellow) are displayed in their respective electron densities for clarity. The inset in (A) is a close-up view of the RMF electron density in the region between Trp^{40} and Trp^{44}.

$I/\sigma I = 2$ are 3.2 Å, 3.4 Å, and 2.9 Å, respectively. The structure of each complex was solved by molecular replacement using, as the starting model, the atomic coordinates of both ribosomal subunits from the published structure of the *Tth* 70*S* ribosome with ligands removed (*13*). For all complexes, the initial unbiased difference electron density maps calculated with the $F_{obs} - F_{calc}$ amplitudes showed positive density corresponding to the known structures of each of the hibernation factors (Fig. 1, A to C, and fig. S1, A, C, and E). The statistics for data processing and refinement of all complexes are shown in table S1 (*14*).

The 3.0 Å resolution crystal structure of RMF bound to the *Tth* 70*S* ribosome reveals that RMF binds next to the 3′ end of the 16*S* ribosomal RNA (rRNA) at the anti–Shine-Dalgarno (anti-SD) region between the head and the platform domains in the mRNA exit channel (Fig. 2, A and B, and movie S1). It interacts with three nucleotides of the 16*S* rRNA that are upstream of the anti-SD sequence (Fig. 2C). This position of RMF is incompatible with the SD of the mRNA making interactions with the anti-SD (Fig. 2D). During the first stages of initiation of protein synthesis, the mRNA forms a double helix with the anti-SD sequence of the 3′ end of the 16*S* rRNA (*15*, *16*). RMF thus prevents initiation of protein synthesis by sterically interfering with the formation of the double helix between the anti-SD and the SD sequence of an incoming mRNA.

The position of RMF in our complex contradicts the conclusions from previous cross-linking (*7*) and chemical-probing experiments (*17*), which were interpreted to suggest that the RMF binding site is located in the vicinity of the peptidyl transferase center (PTC) on the 50*S* subunit (fig. S2, A and B). Our structural data are not consistent with the formation of cross-links between RMF and ribosomal proteins S13, L2, and L13 that were reported. However, no cross-links were observed between RMF and other ribosomal proteins that are located closer to the PTC and are accessible to a cross-linking agent.

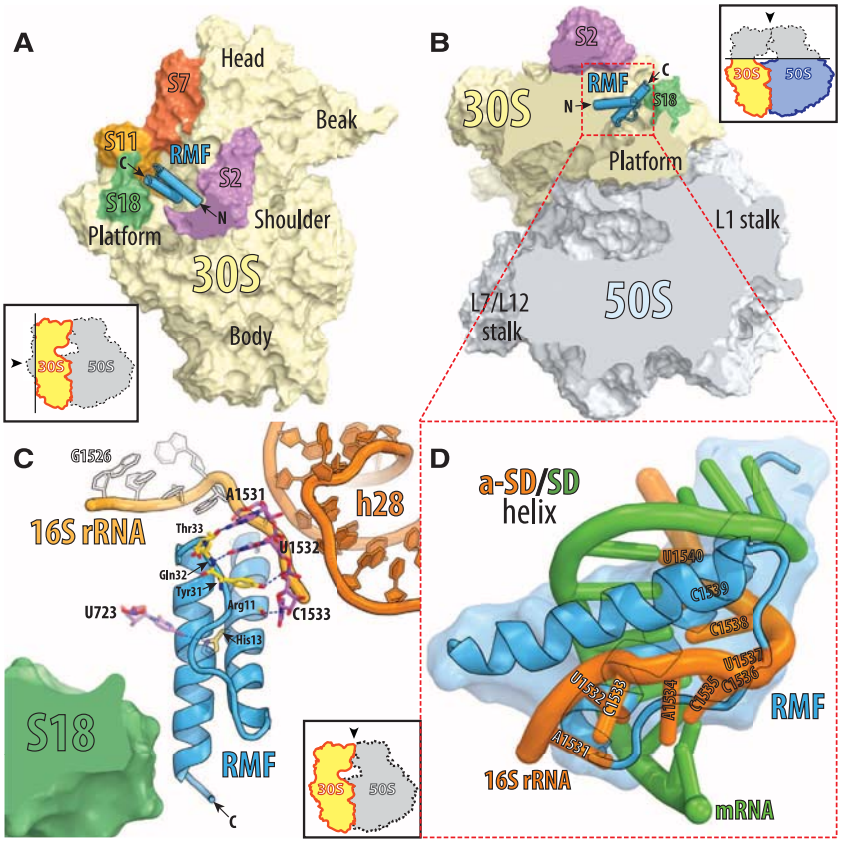


Fig. 2. The structure of RMF bound to the 70*S* ribosome. (**A**) RMF (blue) bound to the 70*S* ribosome viewed from the cytosolic side of the 30*S* subunit (light yellow). As indicated by the inset, the 50*S* subunit and part of ribosomal protein S2 (magenta) are omitted for clarity. The ribosomal proteins surrounding the RMF binding site are colored red for S7, orange for S11, and green for S18. (**B**) The same as (A) but viewed from the top after removing the head of the 30*S* subunit (light yellow) and the protuberances of the 50*S* subunit (light blue) as indicated by the inset. (**C**) Detailed interactions of RMF with components of the 30*S* subunit. Different hues of orange indicate parts of the 16*S* rRNA: light orange for the 3′ end and dark orange for helix 28. C-terminal helix of the RMF comes close enough to ribosomal protein S18 to allow direct contacts, although specific interactions were not identified because of disorder of the S18 side chains. (**D**) A close-up view of the sterical clash between an mRNA (green) involved in the Shine-Dalgarno interactions (*25*) (PDB entry: 2HGR) and ribosome-bound RMF (blue). The anti-SD part of 16*S* rRNA is in orange. Residues 1531 to 1540 of the 16*S* rRNA are indicated.

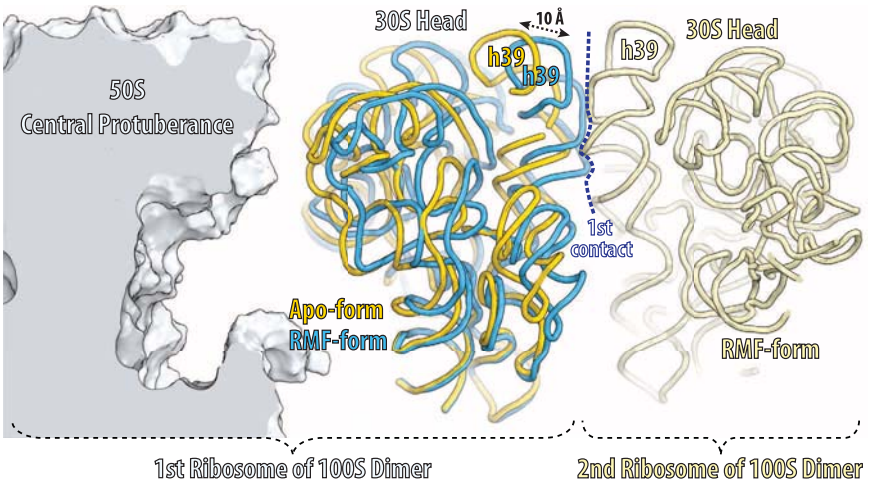


Fig. 3. RMF-induced reorientation of the 30*S*-subunit head domain. The orientation of the two RMF-bound 70*S* monomers in the 100*S* dimer was derived from the published electron density for the 100*S* particle (*9*) (Electron Microscopy Data Bank entry: 1750). Parts of the first 70*S* monomer are displayed: 50*S* subunit (light blue), apo or YfiA-bound form of the 30*S* subunit (yellow); RMF-bound form of the 30*S* subunit (dark blue). Parts of the second 70*S* monomer are displayed in light yellow.

The results from chemical probing suggested that binding of RMF caused conformational changes to several bases in the 50*S* subunit that are located around the PTC (*17*). However, in that study, only regions of 23*S* rRNA closer to the PTC were selected for analysis; the 16*S* rRNA was not analyzed. Therefore, those experiments do not exclude the possibility that RMF does bind to the 30*S* subunit. Moreover, we observed that the 30*S* subunits dimerize in vitro upon binding RMF, which is in agreement with the presumed location of the RMF binding on the 30*S* and not the 50*S* subunit (fig. S3B). Additionally, the only unassigned electron density in both recent cryo-EM reconstructions of 100*S* dimers (*8*, *9*) coincides with the binding site of RMF on the 30*S* subunit, as observed in our structure.

The binding of RMF to the ribosome induces a conformational change (Fig. 3) that facilitates the formation of 100*S* ribosome dimers in which the 30*S* subunits of the two 70*S* ribosomes interact to form a two-fold symmetrical particle (*8*, *9*) with two prominent contact regions: One is centered around ribosomal protein S2, and the other is between the head domains (fig. S4A). The dimerization is not mediated by the direct contact between the RMF molecules, as they are ~100 Å apart in our 100*S* dimer model and are not located at the dimer interface (fig. S4B). Therefore, we hypothesize that RMF promotes the formation of a 100*S* dimer by inducing the movement of the head domains of the 30*S* subunits toward each other (movie S1). In our structure, the head domain of the 30*S* subunit is repositioned away from the central protuberance of the 50*S* subunit to an orientation that increases the contact surface between the head domains, and this increased contact promotes formation of the 100*S* dimer (Fig. 3). The rotation axis of the head domain is perpendicular to the one observed in recent studies of the ratcheted ribosome (*18*) (movie S2). The binding of HPF to the ribosome induces the head domain of the 30*S* subunit to adopt a conformation similar to the one observed with bound RMF, consistent with its role in stabilizing 100*S* dimer formation. Neither the apo-state nor the YfiA-bound state of the ribosome stabilizes a dimerization-competent orientation of the small subumit.

The electron density is observed for RMF bound to both copies of the ribosome in the asymmetric unit. Although the conformation of the head domain in one copy is compatible with 100*S* dimer formation, its orientation in the other copy is similar to the classical state (*13*), presumably because of important crystal contacts made by the head domain in this copy. The equilibrium between the apo- and dimer-forming conformations of the head domain can be altered by a number of factors, e.g., monomeric 70*S* apo-ribosomes can be converted to 100*S* dimers simply by increasing the Mg^{2+} concentration in vitro (*19*). Note that one of the two contacts between the 30*S* subunits in the dimer (fig. S4, A and B) is similar to a crystal contact observed in crystals of the *Tth* 30*S* subunit (*20*).

The observed location of HPF (Fig. 4, A and B, and fig. S5) and YfiA (Fig. 4, C and D, and fig. S5) in our 3.1 Å and 2.75 Å resolution maps, respectively, is in general agreement with the published models derived from an 11.5 Å resolution map of the

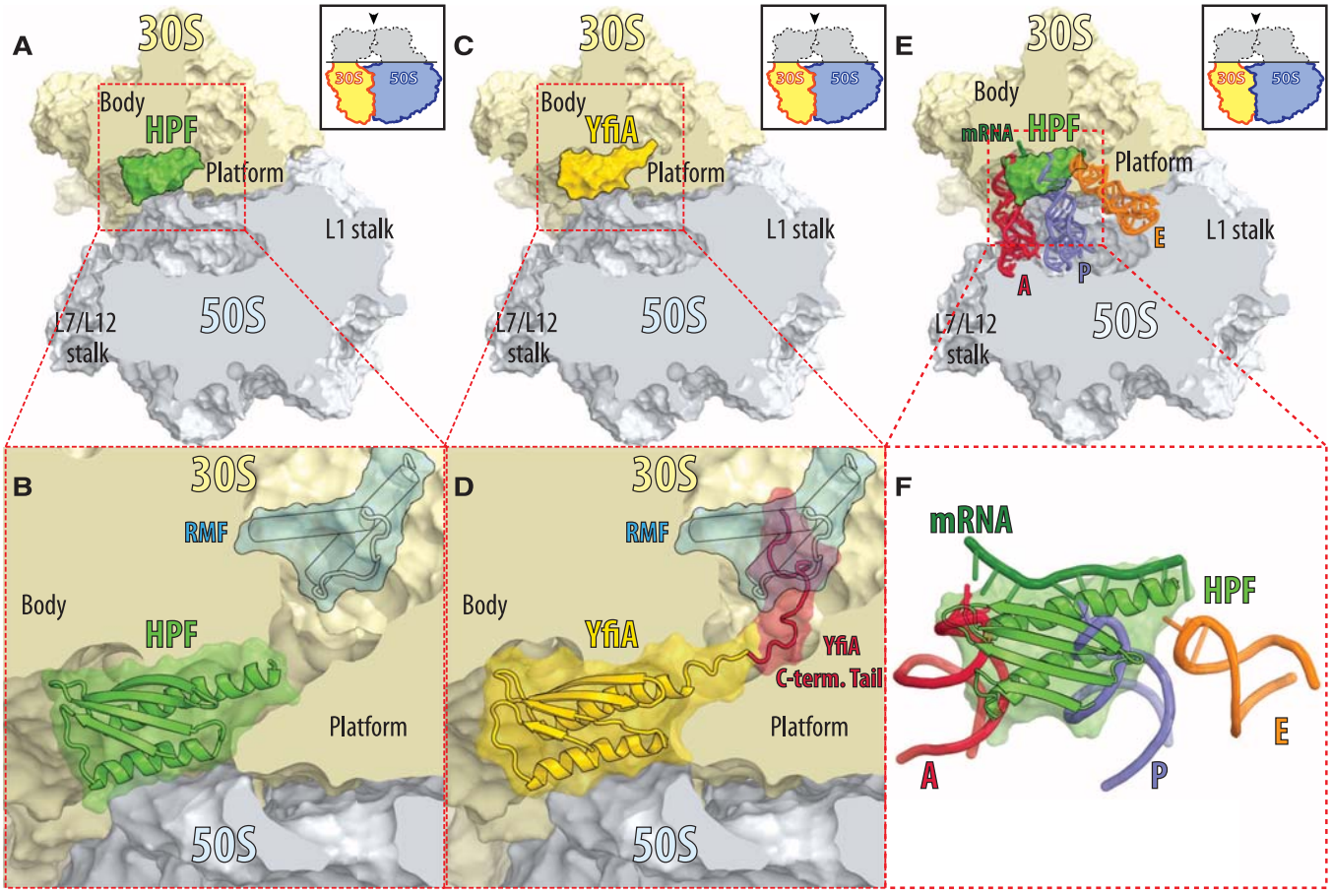


Fig. 4. The structures of HPF and YfiA bound to the 70*S* ribosome. (**A**) HPF (green) bound in the mRNA channel between the head and the body of the 30*S* subunit (light yellow), viewed from the top after removing the head of the 30*S* subunit (light yellow) and the protuberances of the 50*S* subunit (light blue), as indicated by the inset. (**B**) Close-up view of (A) with the RMF protein superimposed (blue). (**C**) YfiA (yellow) bound to the 70*S* ribosome. The view is the same as in (A). (**D**) Close-up view of (C) with the RMF protein superimposed (blue) and the solution structure–derived model of the C-terminal tail of YfiA (red) (*26*). Electron density was observed only for the N-terminal part (residues 90 to 96) of the 24-residue tail; the remaining 17 residues are disordered. (**E**) Steric interference of HPF with the binding of mRNA and tRNAs. The structure of HPF (green) bound to the ribosome with superimposed mRNA (dark green) and tRNAs in the A site (red), P site (blue), and E site (orange) (*13*) (PDB entry: 2J00). View is the same as in (A). (**F**) A close-up view of (E) but the ribosome is omitted for clarity.

analogous complex between YfiA and the *Eco* ribosome (*11*), as well as with 8.5 Å resolution cryo-EM density map of a similar complex between chloroplast-specific YfiA homolog, PSRP1, bound to the *Eco* ribosome (*12*). In our model, HPF and YfiA are bound in the channel that lies between the head and the body of the 30*S* subunit where tRNAs and mRNA bind during protein synthesis (Fig. 4, E and F). Although the globular domains of HPF and YfiA have overlapping binding sites, HPF stabilizes the 100*S* ribosome dimer, whereas YfiA inhibits its formation (*6*). The inhibition is due to the extended C-terminal tail of YfiA, which HPF does not have, that blocks the binding of RMF and, thus, the RMF-induced dimer formation. The visible portion of the YfiA tail follows the mRNA channel, whereas the C-terminal end of the tail, which presumably projects into the RMF binding site (Fig. 4D), could not be modeled. This observation is consistent with previous biochemical studies suggesting mutually exclusive binding of YfiA and RMF to the 70*S* ribosome (*5*). Additionally, YfiA also prevents the formation of the 100*S* dimer by stabilizing the head domain of the small subunit in its apo-conformation.

Our structures reveal that RMF and HPF can bind simultaneously and function together to interfere with the initiation of protein synthesis, which is consistent with the biochemical data (*5*). The HPF-YfiA binding site not only overlaps with all of the tRNA binding sites (Fig. 4, E and F) but also with the binding sites of the initiation factors IF1 and IF3 (fig. S6, A and B), which are directly involved in dissociation of the ribosomes into subunits (*21*), and elongation factor G (fig. S6, C and D), which assists ribosome recycling factor in dissociating posttermination complexes of 70*S* ribosomes (*22*). The inability of these factors to perform their function as a result of blocking by HPF or YfiA of their binding sites explains the reduced dissociation of the stationary-phase ribosomes into subunits (*10*). Because RMF, HPF, and YfiA bind exclusively to the 30*S* subunit, they might not only interfere with initiation of protein synthesis starting on the 70*S* ribosomes, as in the case of leaderless mRNA (*23*) or during reinitiation along polycistronic mRNAs (*24*) but also with canonical initiation starting on the 30*S* subunits.

These studies show that these stationary-phase proteins, when bound to the ribosome, sterically clash with mRNA and tRNAs, and therefore, they cannot act on actively translating ribosomes. This ensures that the stationary-phase factors function only after the completion of the ongoing translation cycles and can only act effectively during stress and/or starvation conditions when the availability of the mRNA and/or tRNAs is limiting.

References and Notes

1. A. Wada, Y. Yamazaki, N. Fujita, A. Ishihama, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 2657 (1990).
2. D. E. Agafonov, V. A. Kolb, A. S. Spirin, *EMBO Rep.* **2**, 399 (2001).
3. J. L. Martinez, F. Baquero, *Antimicrob. Agents Chemother.* **44**, 1771 (2000).
4. M. Kivisaar, *Environ. Microbiol.* **5**, 814 (2003).
5. M. Ueta *et al.*, *Genes Cells* **10**, 1103 (2005).
6. M. Ueta *et al.*, *J. Biochem.* **143**, 425 (2008).
7. H. Yoshida *et al.*, *J. Biochem.* **132**, 983 (2002).
8. T. Kato *et al.*, *Structure* **18**, 719 (2010).
9. J. O. Ortiz *et al.*, *J. Cell Biol.* **190**, 613 (2010).
10. D. E. Agafonov, V. A. Kolb, I. V. Nazimov, A. S. Spirin, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 12345 (1999).
11. A. Vila-Sanjurjo, B. S. Schuwirth, C. W. Hau, J. H. Cate, *Nat. Struct. Mol. Biol.* **11**, 1054 (2004).
12. M. R. Sharma *et al.*, *J. Biol. Chem.* **285**, 4006 (2010).
13. M. Selmer *et al.*, *Science* **313**, 1935 (2006).
14. Materials and methods are available as supplementary materials on *Science* Online.
15. J. Shine, L. Dalgarno, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 1342 (1974).
16. J. A. Steitz, K. Jakes, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 4734 (1975).
17. H. Yoshida, H. Yamamoto, T. Uchiumi, A. Wada, *Genes Cells* **9**, 271 (2004).
18. H. Jin, A. C. Kelley, V. Ramakrishnan, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 15798 (2011).
19. A. Tissières, J. D. Watson, D. Schlessinger, B. R. Hollingworth, *J. Mol. Biol.* **1**, 221 (1959).
20. B. T. Wimberly *et al.*, *Nature* **407**, 327 (2000).
21. M. Y. Pavlov, A. Antoun, M. Lovmar, M. Ehrenberg, *EMBO J.* **27**, 1706 (2008).
22. L. Janosi, H. Hara, S. J. Zhang, A. Kaji, *Adv. Biophys.* **32**, 121 (1996).
23. I. Moll, G. Hirokawa, M. C. Kiel, A. Kaji, U. Bläsi, *Nucleic Acids Res.* **32**, 3354 (2004).
24. A. L. Karamyshev, Z. N. Karamysheva, T. Yamami, K. Ito, Y. Nakamura, *Biochimie* **86**, 933 (2004).
25. G. Yusupova, L. Jenner, B. Rees, D. Moras, M. Yusupov, *Nature* **444**, 391 (2006).
26. A. Rak, A. Kalinin, D. Shcherbakov, P. Bayer, *Biochem. Biophys. Res. Commun.* **299**, 710 (2002).

Acknowledgments: We thank the staff at the National Synchrotron Light Source (beamlines X29 and X25) and at the Advanced Photon Source (beamline 24ID) for help during data collection, the staff at the Center for Structural Biology at Yale University for computational support, and S. Seetharaman and M. Gagnon for valuable discussions. This work was supported by NIH grant GM022778 awarded to T.A.S. The structure factors and coordinates for both copies of the 70*S* ribosome in the asymmetric unit of all complexes have been deposited in the Research Collaboration for Structural Biology Protein Data Bank with the following accession codes: 3V22, 3V23, 3V24, and 3V25 for the RMF-ribosome complex; 3V26, 3V27, 3V28, and 3V29 for the HPF-ribosome complex; and 3V2C, 3V2D, 3V2F, and 3V2E for the YfiA-ribosome complex. T.A.S. owns stock in and is on the advisory board of Rib-X Pharmaceuticals, Inc., which does structure-based drug design targeted at the ribosome.

Supplementary Materials

www.sciencemag.org/cgi/content/full/336/6083/915/DC1
Materials and Methods
Figs. S1 to S6
Table S1
References (*27*–*41*)
Movies S1 and S2

29 December 2011; accepted 13 April 2012
10.1126/science.1218538

The Ancient Drug Salicylate Directly Activates AMP-Activated Protein Kinase

Simon A. Hawley,[1] Morgan D. Fullerton,[2] Fiona A. Ross,[1] Jonathan D. Schertzer,[2] Cyrille Chevtzoff,[1] Katherine J. Walker,[1] Mark W. Peggie,[3] Darya Zibrova,[3] Kevin A. Green,[1] Kirsty J. Mustard,[1] Bruce E. Kemp,[4] Kei Sakamoto,[3]* Gregory R. Steinberg,[2,4] D. Grahame Hardie[1]†

Salicylate, a plant product, has been in medicinal use since ancient times. More recently, it has been replaced by synthetic derivatives such as aspirin and salsalate, both of which are rapidly broken down to salicylate in vivo. At concentrations reached in plasma after administration of salsalate or of aspirin at high doses, salicylate activates adenosine monophosphate–activated protein kinase (AMPK), a central regulator of cell growth and metabolism. Salicylate binds at the same site as the synthetic activator A-769662 to cause allosteric activation and inhibition of dephosphorylation of the activating phosphorylation site, threonine-172. In AMPK knockout mice, effects of salicylate to increase fat utilization and to lower plasma fatty acids in vivo were lost. Our results suggest that AMPK activation could explain some beneficial effects of salsalate and aspirin in humans.

The medicinal effects of willow bark have been known since the time of Hippocrates. The active component is salicylate, a hormone produced by plants in response to pathogen infection (*1*). For medicinal use, it was largely replaced by aspirin (acetyl salicylate), which is rapidly broken down to salicylate in vivo (*2*, *3*). Salicylate can also be administered as salsalate, which shows promise for treatment of insulin resistance and type 2 diabetes (*4*, *5*). Aspirin and salicylate inhibit cyclo-oxygenases and, hence, prostanoid biosynthesis (*6*), as well as the protein kinase IκB kinase β (IKKβ) in the NF-κB pathway (*7*). However, some effects of these drugs are still observed in mice deficient in these pathways (*8*).

Adenosine monophosphate–activated protein kinase (AMPK) is a cellular energy sensor con-

[1]Division of Cell Signalling and Immunology, College of Life Sciences, University of Dundee, Dundee DD1 5EH, Scotland, UK. [2]Divisions of Endocrinology and Metabolism, Department of Medicine, and Department of Biochemistry and Biomedical Sciences, McMaster University, 1200 Main Street West, Hamilton, Ontario L8N 3Z5, Canada. [3]Medical Research Council Protein Phosphorylation Unit, College of Life Sciences, University of Dundee, Dundee DD1 5EH, Scotland, UK. [4]St. Vincent's Institute of Medical Research and Department of Medicine, University of Melbourne, 41 Victoria Parade, Fitzroy, Vic 3065 Australia.

*Present address: Nestlé Institute of Health Sciences Société Anonyme, Campus École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland.
†To whom correspondence should be addressed: E-mail: d.g.hardie@dundee.ac.uk

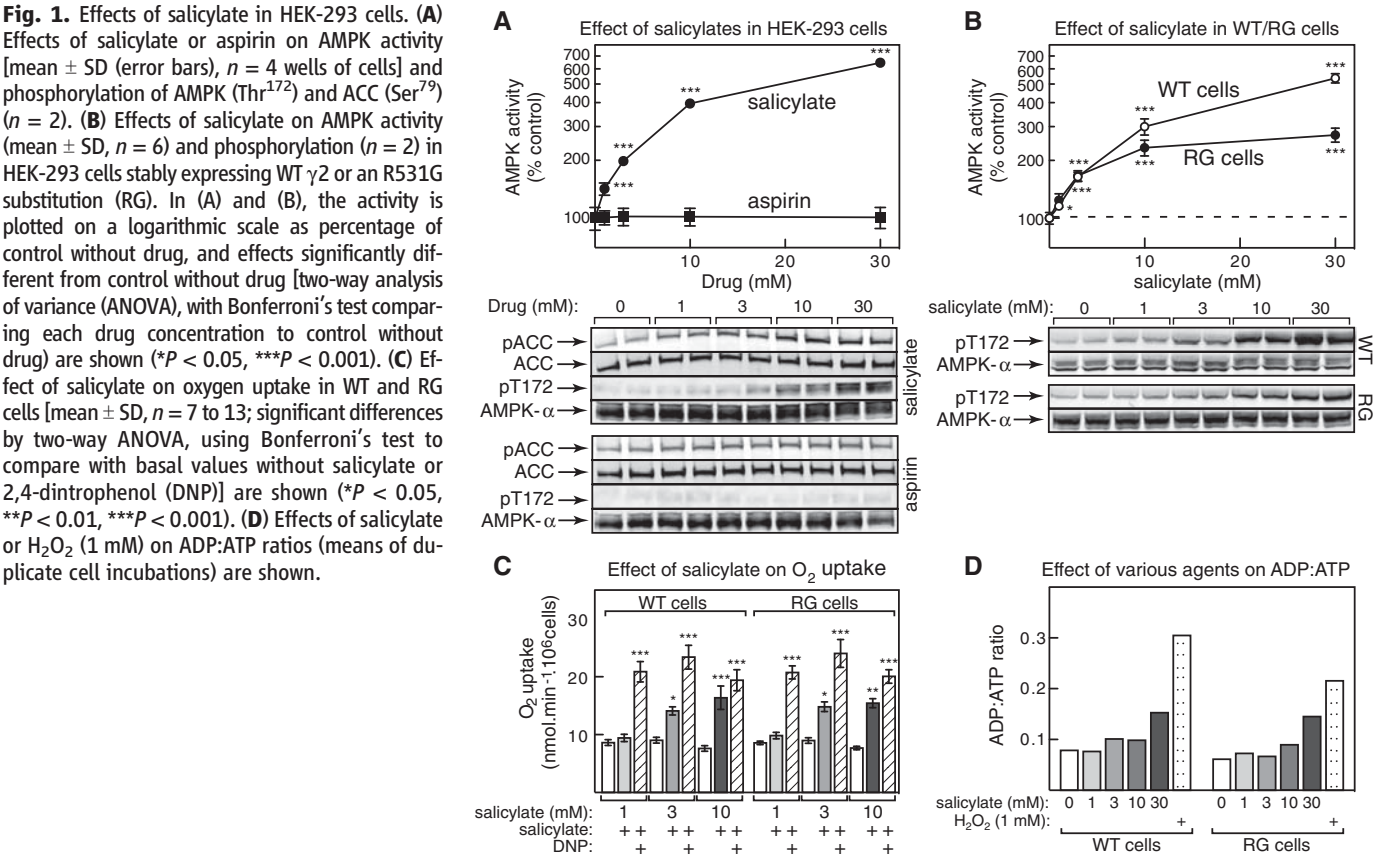


Fig. 1. Effects of salicylate in HEK-293 cells. (**A**) Effects of salicylate or aspirin on AMPK activity [mean ± SD (error bars), $n = 4$ wells of cells] and phosphorylation of AMPK (Thr^{172}) and ACC (Ser^{79}) ($n = 2$). (**B**) Effects of salicylate on AMPK activity (mean ± SD, $n = 6$) and phosphorylation ($n = 2$) in HEK-293 cells stably expressing WT γ2 or an R531G substitution (RG). In (A) and (B), the activity is plotted on a logarithmic scale as percentage of control without drug, and effects significantly different from control without drug [two-way analysis of variance (ANOVA), with Bonferroni's test comparing each drug concentration to control without drug] are shown (*$P < 0.05$, ***$P < 0.001$). (**C**) Effect of salicylate on oxygen uptake in WT and RG cells [mean ± SD, $n = 7$ to 13; significant differences by two-way ANOVA, using Bonferroni's test to compare with basal values without salicylate or 2,4-dintrophenol (DNP)] are shown (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$). (**D**) Effects of salicylate or H_2O_2 (1 mM) on ADP:ATP ratios (means of duplicate cell incubations) are shown.

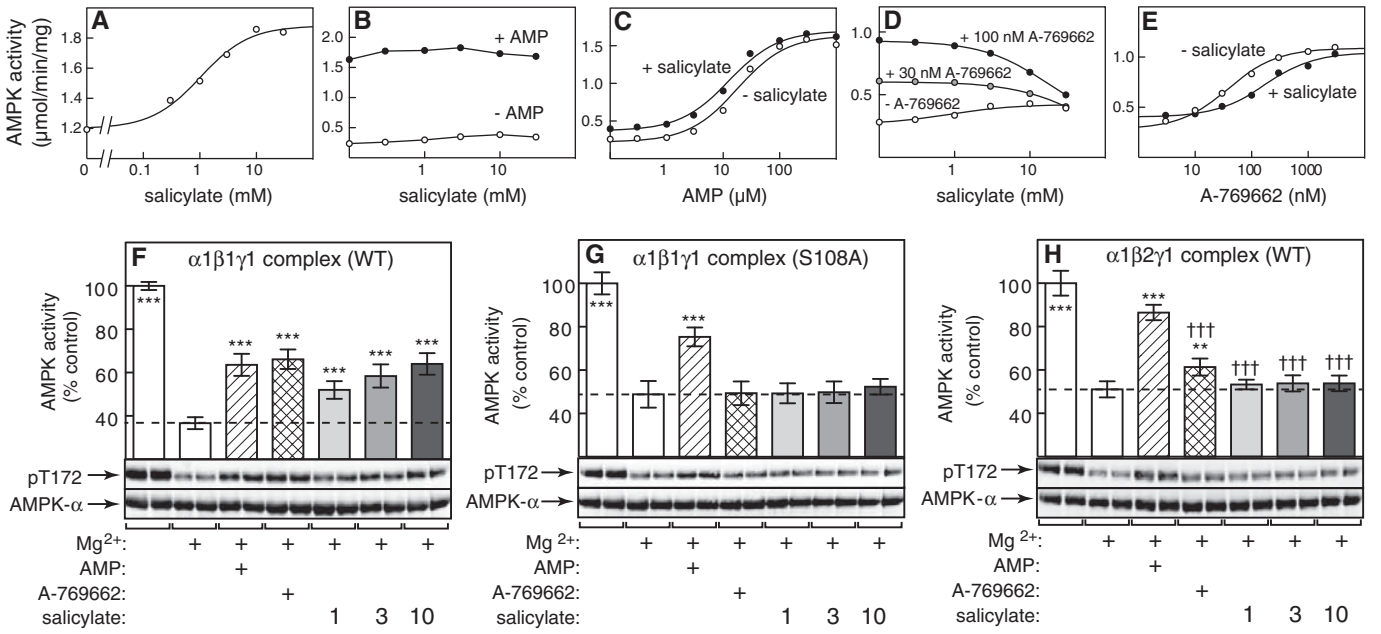


Fig. 2. Effect of salicylate on AMPK in cell-free assays. (**A** to **E**) Effects of salicylate on activity of purified rat liver AMPK: (A) effect of salicylate, (B) effect of salicylate ± 200 μM AMP, (C) effect of AMP ± 10 mM salicylate, (D) effect of salicylate ± 30 and 100 nM A-769662, and (E) effect of A-769662 ± 10 mM salicylate. Data points in (A) to (E) are means of duplicate assays; lines were generated by fitting data to the equation: activity = basal + [basal*(activation − basal)*$X/(A_{0.5} + X)$], where X is the concentration of the activator. Values obtained for $A_{0.5}$ and activation are quoted in the text. (**F** to **H**) Effects of AMP, A-769662, and salicylate on dephosphorylation of bacterially expressed human AMPK complexes by PP2Cα (all incubations contained PP2Cα, but control lacked Mg^{2+}); bar graphs show AMPK activity (percentage of control without Mg^{2+}, mean ± SD, $n = 6$); gel pictures show Thr^{172} phosphorylation ($n = 2$). (F) WT α1β1γ1 complex, (G) α1β1γ1 complex with β1 S108A substitution, and (H) WT α1β2γ1 complex. Significant differences from the control plus Mg^{2+}, using one-way ANOVA with Dunnett's multiple comparison test, are shown: **$P < 0.01$, ***$P < 0.001$. Also shown are significant differences in the size of the effect of A-76962 or salicylate between the α1β1γ1 and α1β2γ1 complexes (2F and 2H): †††$P < 0.001$. For the latter comparisons, the differences between the $+Mg^{2+}$+A-76922 or $+Mg^{2+}$ +salicylate and the $+Mg^{2+}$ only columns were first expressed as a fraction of the difference between the $+Mg^{2+}$+AMP and $+Mg^{2+}$ only columns.

served throughout eukaryotes. This heterotrimeric enzyme is composed of catalytic α subunits and regulatory β and γ subunits (*9*, *10*). Once activated in response to metabolic stress, AMPK phosphorylates targets that switch off adenosine triphosphate (ATP)–consuming processes while switching on catabolic pathways that generate ATP. AMPK is activated >100-fold by phosphorylation at Thr^{172} in the α subunit by the tumor suppressor protein kinase, LKB1, or the Ca^{2+}-dependent kinase, CaMKKβ (calmodulin-dependent kinase kinase-β) (*9*, *10*). Binding of AMP or adenosine diphosphate (ADP) to the γ subunit triggers a conformational change that promotes phosphorylation and inhibits dephosphorylation (*11*–*15*), causing a switch to the active form. Binding of AMP (but not ADP) to a second site (*15*) causes further allosteric activation, leading to >1000-fold activation overall (*16*). Most drugs or xenobiotics that activate AMPK work by inhibiting mitochondrial ATP synthesis and increasing the concentration of AMP and ADP (*17*). However, a synthetic activator, A-769662 (*18*), which also causes allosteric activation and inhibits Thr^{172} dephosphorylation, binds directly to AMPK at sites distinct from those used by AMP (*19*–*21*).

Salicylate, but not aspirin, activated AMPK when applied to human embryonic kidney (HEK) 293 cells, with its effects being significant at 1 mM and above (Fig. 1A; it appears that the esterases that catalyze breakdown of aspirin to salicylate are not expressed in these cells). This was associated with increased phosphorylation of Thr^{172} on AMPK-α and the downstream target of AMPK, acetyl–coenzyme A carboxylase (ACC); in the latter case, the effects were evident at 1 mM and above (Fig. 1A). Salicylate can uncouple mitochondrial respiration (*22*), so we suspected that it might activate AMPK by decreasing cellular ATP and increasing AMP and ADP. To test this, we used isogenic cell lines expressing wild-type AMPK (WT cells) or a mutated enzyme in which an Arg^{531}→Gly^{531} (R531G) substitution in γ2 renders AMPK insensitive to AMP or ADP (RG cells) (see supplemental materials and methods) (*15*, *17*). At concentrations <10 mM, salicylate caused similar increases in AMPK phosphorylation or activation in WT and RG cells, showing that the effect was not dependent on changes in AMP or ADP. However, at 10 mM and above, there was a greater activation/phosphorylation in WT than in RG cells, suggesting that AMP- or ADP-dependent effects were also occurring at these higher concentrations (Fig. 1B). Concentrations >1 mM salicylate increased cellular oxygen uptake in both WT and RG cells. This effect was not additive with the effect of a concentration of dinitrophenol causing a maximal increase in oxygen uptake, suggesting that at these concentrations, salicylate, like dinitrophenol, could dissipate the proton gradient and thus uncouple the respiratory chain from ATP synthesis (Fig. 1C). However, unlike effects of another AMPK activator (H_2O_2), any increases in the cellular ADP:ATP ratio at salicylate concentrations below 30 mM were very small (Fig. 1D). Thus, mitochondria appear to compensate for mild uncoupling by increasing respiration. Salicylate does not activate AMPK through the Ca^{2+}-CaMKKβ pathway (*12*), because the CaMKK inhibitor STO-609 had no effect on responses to salicylate, although it blocked responses to a Ca^{2+} ionophore, A23187 (fig. S1, A and B, supplementary data).

Using a physiological concentration of ATP in assays (2 mM), salicylate caused a 1.6-fold allosteric activation with a half-maximal effect ($A_{0.5}$) at 1.0 ± 0.2 mM (Fig. 2A). We observed a large activation by AMP at all salicylate concentrations

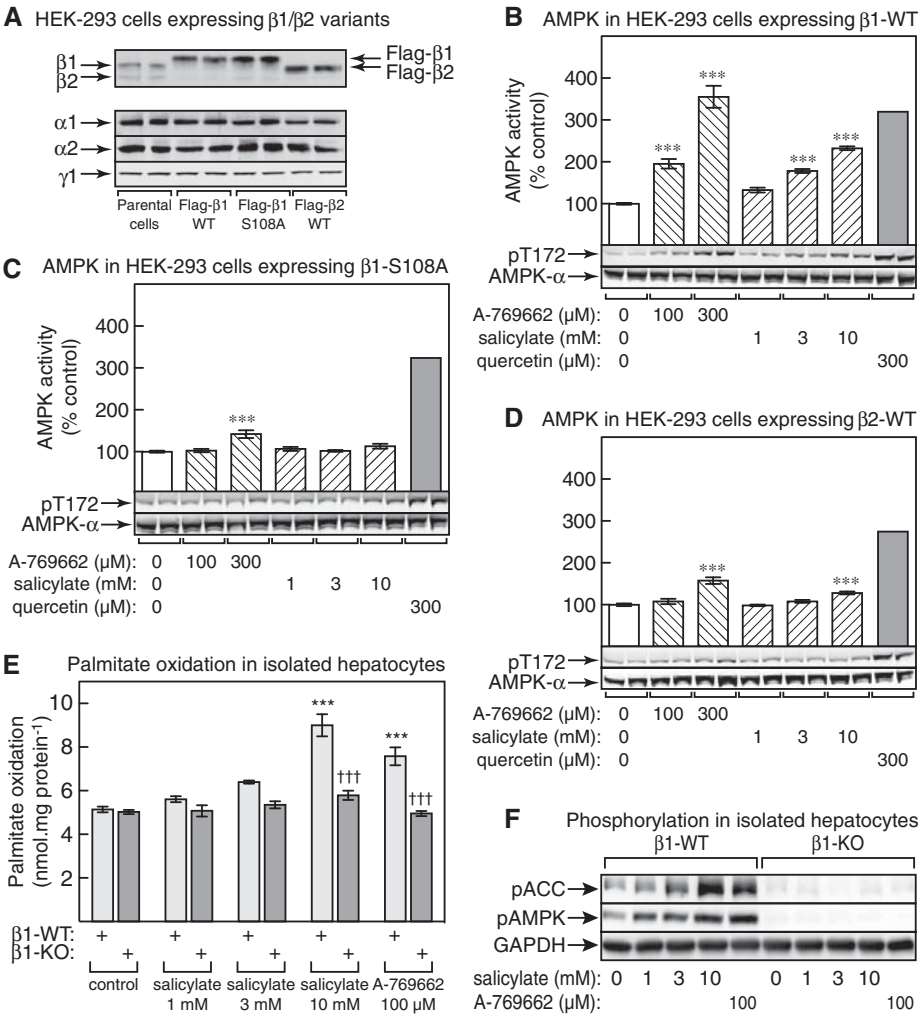


Fig. 3. Effects of salicylate and A-769662 in intact cells. (**A**) Expression of β subunits assessed using pan-β antibody in parental cells or cells stably expressing β1 WT, β1-S108A, or β2 WT and of endogenous α1, α2, and γ1 in the same cells. (**B** to **D**) Activity and phosphorylation of AMPK after treatment with various activators in cells expressing (B) β1 WT, (C) β1-S108A, and (D) β2 WT. Kinase assays [mean ± SEM, $n = 6$ except for 100 μM A-769662 ($n = 4$) and quercetin ($n = 2$); significantly different from control without drug, by one-way ANOVA with Dunnett's multiple comparison test, ***$P < 0.001$] and Western blots ($n = 2$) were of immunoprecipitates made using an antibody to FLAG. (**E**) Palmitate oxidation in hepatocytes isolated from β1-KO mice and WT controls (mean ± SEM, $n = 6$ to 14, significantly different from control without drug by two-way ANOVA with Bonferroni's test, ***$P < 0.001$; †††significantly different from WT, $P < 0.001$). (**F**) Phosphorylation of AMPK and ACC in hepatocytes isolated from β1-KO or WT mice. GAPDH, glyceraldehyde-3-phosphate dehydrogenase.

up to 30 mM (Fig. 2B), and salicylate did not affect the concentration of AMP causing half-maximal activation ($A_{0.5}$ = 18 ± 3 or 13 ± 2 μM, with or without 10 mM salicylate) (Fig. 2C). In contrast, increasing concentrations of salicylate progressively antagonized activation by 30 and 100 nM A-769662 (Fig. 2D), and 10 mM salicylate increased $A_{0.5}$ for A-769662 by >fourfold (from 39 ± 4 to 172 ± 24 nM) (Fig. 2E). These results suggest that salicylate is a partial agonist acting at the same site as A-769662, causing a small activation on its own but antagonizing the larger activation by A-769662. If it binds at the same site, salicylate should protect against Thr^{172} dephosphorylation, as does A-769662 (*19*, *20*). Indeed, salicylate protected AMPK (human α1β1γ1 complex) against dephosphorylation and inactivation by protein phosphatase-2Cα to the same extent as did AMP and A-769662 (Fig. 2F), although it had no effect on PP2Cα assayed using a peptide substrate (fig. S2). A Ser^{108}→Ala^{108} (S108A) substitution in the β1 subunit abolishes effects of A-769662 on dephosphorylation, and β2-containing complexes are resistant to the drug (*20*, *21*). As expected, an S108A substitution in the α1β1γ1 complex abolished the effects of A-769662 and salicylate, but not that of AMP (Fig. 2G), whereas there was no effect of salicylate on dephosphorylation of an α1β2γ1 complex, although there was a small effect of A-769662 (Fig. 2H). These results lend further support to the idea that salicylate binds at the same site(s) as A-769662.

To examine effects of salicylate on AMPK phosphorylation and activation in intact cells, we used HEK-293 cells carrying a Flp recombinase target site to generate isogenic lines expressing FLAG-tagged WT β1, a β1-S108A mutant, or WT β2. With the use of an antibody that recognizes both isoforms, endogenous β1 and β2 could be detected in the parental cells, but, as observed previously when expressing α subunits using this system (*17*), these were largely replaced by FLAG-tagged β1 and β2 (with reduced electrophoretic mobility) in cells expressing recombinant β1/β2; the expression of α1, α2, and γ1 subunits was unaffected (Fig. 3A). A-769662 and salicylate caused increased activation/phosphorylation of AMPK in cells expressing WT β1 (Fig. 3B), but their effects were greatly reduced in cells expressing the β1-S108A mutant (Fig. 3C) or WT β2 (Fig. 3D), In contrast, the effects of quercetin, which acts by increasing AMP (*17*), were unaffected.

We used β1 knockout (KO) mice to test whether salicylate has metabolic effects in vivo through AMPK (*23*). Fatty-acid oxidation in isolated WT hepatocytes was stimulated by salicylate or A-769662 and was associated with increased phosphorylation of AMPK and ACC [the latter regulating fat oxidation (*24*)]. All effects were reduced or eliminated in hepatocytes from β1-KOs (Fig. 3, E and F, and fig. S3). There were no changes in ADP:ATP ratios in WT cells in response to 1 to 10 mM salicylate (fig. S4A).

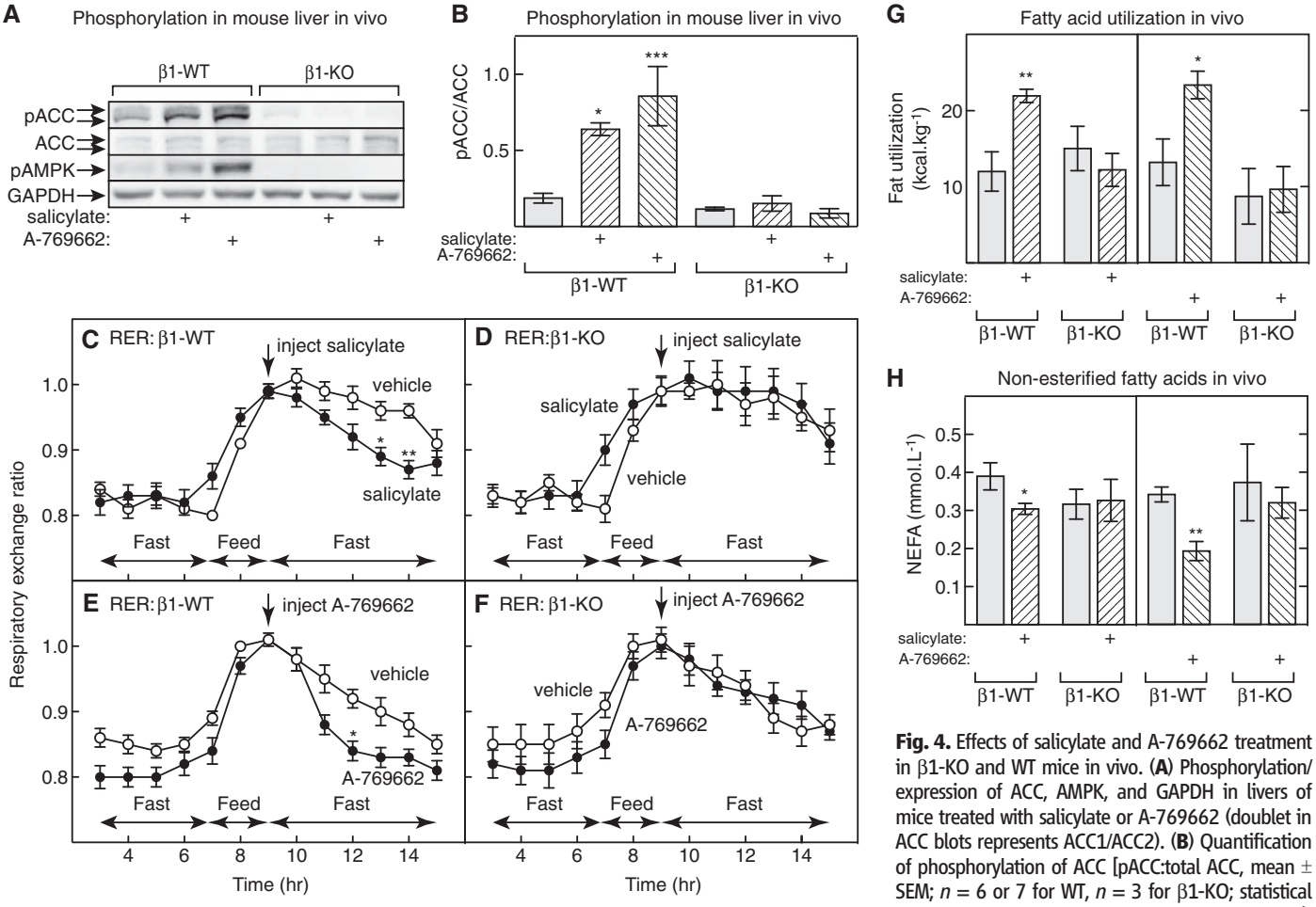


Fig. 4. Effects of salicylate and A-769662 treatment in β1-KO and WT mice in vivo. (**A**) Phosphorylation/expression of ACC, AMPK, and GAPDH in livers of mice treated with salicylate or A-769662 (doublet in ACC blots represents ACC1/ACC2). (**B**) Quantification of phosphorylation of ACC [pACC:total ACC, mean ± SEM; n = 6 or 7 for WT, n = 3 for β1-KO; statistical significance by one-way ANOVA with Bonferroni's test compared with vehicle only is shown (*P < 0.05, ***P < 0.001)]. (**C** to **F**) RER measured in β1-KO and WT mice after injection of vehicle, salicylate (250 mg/kg), or A-769662 (30 mg/kg) at the start of a period of fasting. Results are mean ± SEM (n = 6 to 13). By two-way ANOVA, effects of salicylate (P < 0.05) or A-769662 (P < 0.001) were only significant in WT mice; significant differences by Bonferroni's test at individual time points are shown (*P < 0.05, **P < 0.01). (**G**) Fatty-acid utilization calculated from data in (C) to (F). Results are mean ± SEM (n = 7 or 8). Significant differences by two-way ANOVA with Bonferroni's test are shown (*P < 0.05, **P < 0.01). (**H**) Plasma nonesterified fatty acids (NEFA) in mice treated with salicylate or A-769662 for 90 min. Results are mean ± SEM (n = 7 or 8). Significant differences by two-way ANOVA with Bonferroni's test are shown (*P < 0.05, **P < 0.01).

We injected mice with salicylate or A-769662 at the start of a period of fasting to assess metabolism in vivo. The dose that we used generated plasma salicylate of 2.4 ± 0.4 and 2.0 ± 0.1 mM (mean ± SEM, $n = 7$ and 3 mice) in WT and β1-KO mice, respectively. Both agents caused phosphorylation of liver AMPK and ACC in WT but not β1-KO mice (Fig. 4, A and B), although there were no changes in hepatic AMP:ATP or ADP:ATP ratios in WT mice (fig. S4B). In WT mice, salicylate also caused phosphorylation and activation of AMPK in soleus muscle and adipose tissue (fig. S5). Both agents depressed the respiratory exchange ratio (RER) for 6 hours after injection, consistent with a switch from carbohydrate to fat utilization; these effects were lost in β1-KOs (Fig. 4, C to F). Significant depressions in RER, in WT mice only, were evident by calculating the area under the curve (fig. S6, A and B). When we calculated fat and carbohydrate utilization, both salicylate and A-769662 increased fat utilization in WT but not β1-KO mice (Fig. 4G); A-769662 also decreased carbohydrate utilization in WT mice (fig. S6, C and D). Both salicylate and A-769662 reduced serum nonesterified fatty acids in WT but not β1-KO mice (Fig. 4H). We also studied glucose homeostasis in mice made insulin-resistant by high-fat feeding, followed by daily salicylate injections for 2 weeks. However, effects of salicylate to improve fasting glucose, fasting insulin, glucose tolerance, and insulin resistance (homeostasis model assessment) were retained in β1-KO mice (fig. S7), indicating that they were independent of AMPK.

Our results show that salicylate can directly activate AMPK, primarily by inhibiting Thr^{172} dephosphorylation. The plasma salicylate concentrations in humans treated with oral salsalate (*4*) or high-dose aspirin (30 to 90 mg/kg) (*25*, *26*) are 1 to 3 mM. At these concentrations, salicylate activated AMPK in WT and RG HEK-293 cells to the same extent and did not increase cellular ADP:ATP ratios, indicating an AMP-independent mechanism. Thus, the natural product salicylate can activate AMPK via a mechanism closely related to that of A-769662, a synthetic activator derived from a high-throughput screen [which, unlike salicylates, has poor oral availability (*18*)]. Although the exact site(s) occupied by salicylate and A-769662 on AMPK remain unidentified, our finding that the S108A mutation abolishes activation by both agents suggests that the binding sites overlap.

Effects of salicylate on fat oxidation in vivo appear to require activation of AMPK-β1 complexes. Aspirin also reduces circulating lipids in obese rats and improves insulin sensitivity (*7*). However, in agreement with previous studies (*7*, *27*), our results using long-term salicylate treatment of fat-fed mice (fig. S7) indicate that effects on AMPK-independent pathways, such as IKKβ or c-Jun N-terminal kinase, are also important.

After oral administration, aspirin is rapidly broken down by liver, erythrocyte, and plasma esterases to salicylate (*3*), whose peak plasma concentrations and half-life are orders of magnitude greater than those of aspirin (*2*). Our findings raise the possibility that other effects of aspirin, like protective effects against development of cancer (*28*), may be mediated in part by AMPK. AMPK is activated by the antidiabetic drug metformin (*17*, *29*), and treatment of diabetics with metformin is also associated with reduced cancer incidence (*30*). Our results show that one thing salicylates and metformin have in common is their ability to activate AMPK. However, one caveat is that the doses of aspirin required to activate AMPK in vivo may be higher than those used in most human studies.

Note added in proof: While this paper was being reviewed, Din *et al.* [F. V. N. Din *et al.*, *Gastroenterology* 10.1053/j.gastro.2012.02.050 (2012)] reported that aspirin activates AMPK in colorectal cancer cell lines.

References and Notes

1. P. Reymond, E. E. Farmer, *Curr. Opin. Plant Biol.* **1**, 404 (1998).
2. G. A. Higgs, J. A. Salmon, B. Henderson, J. R. Vane, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 1417 (1987).
3. F. M. Williams, *Clin. Pharmacokinet.* **10**, 392 (1985).
4. A. Fleischman, S. E. Shoelson, R. Bernier, A. B. Goldfine, *Diabetes Care* **31**, 289 (2008).
5. A. B. Goldfine *et al.*, *Ann. Intern. Med.* **152**, 346 (2010).
6. G. A. Higgs, S. Moncada, J. R. Vane, *Ann. Clin. Res.* **16**, 287 (1984).
7. M. Yuan *et al.*, *Science* **293**, 1673 (2001).
8. B. N. Cronstein, M. C. Montesinos, G. Weissmann, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 6377 (1999).
9. D. G. Hardie, *Nat. Rev. Mol. Cell Biol.* **8**, 774 (2007).
10. G. R. Steinberg, B. E. Kemp, *Physiol. Rev.* **89**, 1025 (2009).
11. J. S. Oakhill *et al.*, *Science* **332**, 1433 (2011).
12. S. A. Hawley *et al.*, *J. Biol. Chem.* **270**, 27186 (1995).
13. J. S. Oakhill *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 19237 (2010).
14. S. P. Davies, N. R. Helps, P. T. W. Cohen, D. G. Hardie, *FEBS Lett.* **377**, 421 (1995).
15. B. Xiao *et al.*, *Nature* **472**, 230 (2011).
16. M. Suter *et al.*, *J. Biol. Chem.* **281**, 32207 (2006).
17. S. A. Hawley *et al.*, *Cell Metab.* **11**, 554 (2010).
18. B. Cool *et al.*, *Cell Metab.* **3**, 403 (2006).
19. O. Göransson *et al.*, *J. Biol. Chem.* **282**, 32549 (2007).
20. M. J. Sanders *et al.*, *J. Biol. Chem.* **282**, 32539 (2007).
21. J. W. Scott *et al.*, *Chem. Biol.* **15**, 1220 (2008).
22. T. M. Brody, *J. Pharmacol. Exp. Ther.* **117**, 39 (1956).
23. N. Dzamko *et al.*, *J. Biol. Chem.* **285**, 115 (2010).
24. L. Abu-Elheiga, M. M. Matzuk, K. A. H. Abo-Hashema, S. J. Wakil, *Science* **291**, 2613 (2001).
25. R. O. Day *et al.*, *Br. J. Clin. Pharmacol.* **28**, 695 (1989).
26. R. S. Hundal *et al.*, *J. Clin. Invest.* **109**, 1321 (2002).
27. Z. Gao, A. Zuberi, M. J. Quon, Z. Dong, J. Ye, *J. Biol. Chem.* **278**, 24944 (2003).
28. P. M. Rothwell *et al.*, *Lancet* **377**, 31 (2011).
29. G. Zhou *et al.*, *J. Clin. Invest.* **108**, 1167 (2001).
30. A. A. Ogunleye, S. A. Ogston, A. D. Morris, J. M. Evans, *Br. J. Cancer* **101**, 1199 (2009).

Acknowledgments: D.G.H. was supported by a Programme Grant (080982) from the Wellcome Trust and K.S. by the Medical Research Council. D.G.H. and K.S. were also funded by the companies (AstraZeneca, Boehringer-Ingelheim, GlaxoSmithKline, Merck Serono, and Pfizer) supporting the Division of Signal Transduction Therapy, Univ. of Dundee. The studies were also supported by grants and fellowship from the Canadian Institutes of Health Research (CIHR) (M.D.F. and G.R.S.), the Canadian Diabetes Association (J.D.S. and G.R.S.), the National Health and Medical Research Council of Australia (G.R.S. and B.E.K.), and the Australian Research Council and the Victorian Government's Operational Infrastructure Support Program (B.E.K.). J.D.S. is a DeGroote Academic Fellow (McMaster Univ.), and M.D.F. a Banting Postdoctoral Fellow (CIHR). G.R.S. is a Canada Research Chair in Metabolism and Obesity.

Supplementary Materials

www.sciencemag.org/cgi/content/full/science.1215327/DC1
Materials and Methods
Figs. S1 to S7
References

17 October 2011; accepted 23 March 2012
Published online 19 April 2012;
10.1126/science.1215327

Aerobic Microbial Respiration in 86-Million-Year-Old Deep-Sea Red Clay

Hans Røy,[1]* Jens Kallmeyer,[2] Rishi Ram Adhikari,[2] Robert Pockalny,[3] Bo Barker Jørgensen,[1] Steven D'Hondt[3]

Microbial communities can subsist at depth in marine sediments without fresh supply of organic matter for millions of years. At threshold sedimentation rates of 1 millimeter per 1000 years, the low rates of microbial community metabolism in the North Pacific Gyre allow sediments to remain oxygenated tens of meters below the sea floor. We found that the oxygen respiration rates dropped from 10 micromoles of O_2 $liter^{-1}$ $year^{-1}$ near the sediment-water interface to 0.001 micromoles of O_2 $liter^{-1}$ $year^{-1}$ at 30-meter depth within 86 million-year-old sediment. The cell-specific respiration rate decreased with depth but stabilized at around 10^{-3} femtomoles of O_2 $cell^{-1}$ day^{-1} 10 meters below the seafloor. This result indicated that the community size is controlled by the rate of carbon oxidation and thereby by the low available energy flux.

The discovery of living microbial communities in deeply buried marine sediments (*1*, *2*) has spurred interest in life under extreme energy limitation (*3*). The subtropical gyres are the most oligotrophic regions of the oceans. Primary productivity in the surface waters of the gyres is low, yet within the same order of magnitude as the surrounding open ocean (Fig. 1). Oxygen penetrates many meters into the seabed below the gyres, which indicates extremely low rates of microbial community respiration (*4*, *5*) in contrast to the rest of the seabed

(*4*), where in general oxygen penetration is limited to millimeters to decimeters depth, according to a square root function of the organic matter flux (*6*–*9*).

On R/V *Knorr* voyage 195, we collected sediment cores up to 28 m along the equator and into the North Pacific Gyre and measured the oxygen distribution throughout the retrieved cores by using needle-shaped optical O_2 sensors (PreSens Precision Sensing GmbH, Regensburg, Germany) (Fig. 1). Along the equator, from the Galapagos (site 3) and 4700 km westward into the Pacific Ocean (site 8), the oxygen flux across the sediment-water interface decreased moderately from 60 to 45 mmol m^{-2} $year^{-1}$, whereas the oxygen penetration depth increased from 6 to 9.5 cm (Fig. 1 and table S1). The trend in oxygen flux followed the trend in primary production in the surface water. The increase in oxygen penetration depth in the sediment along the equator is less pronounced than the decrease in oxygen flux, which is expected from the general relation between oxygen flux and oxygen penetration (*10*). From the equator (site 8) and into the North Pacific Gyre (site 11), the primary productivity decreased by 50%, but the oxygen penetration depth increased from 9 cm to greater than 30 m. Such a large change in oxygen penetration cannot be explained by less productivity in the gyre nor by the increase in water depth (*11*).

The volumetric oxygen consumption rates down the length of the core were modeled at high resolution from the deep oxygen profiles we obtained. In the model, we calculated carbon mineralization as the product of the measured particulate organic carbon concentration at each depth (C_{org}) and an initially unknown and depth-dependent reactivity (k) of the C_{org}. For simplicity, we assume a 1:1 molar ratio between carbon oxidation and O_2 consumption. We further assumed that k decreases monotonically with time (t), and that the decrease can be described by a power law function (*12*):

$$k = A \times (t + t_0)^B \quad (1)$$

The mass balance of oxygen throughout the core, taking into account molecular diffusion and oxygen consumption, was solved using the software Comsol Multiphysics. Measured depth-dependent values for porosity, molecular diffusion coefficient (D_m), and C_{org} were fed into the model as smoothed tabulated files and interpolated to fit the modeling grid. Parameters in Eq. 1 were varied to find the best fit to the measured deep oxygen profiles, resulting in the volumetric oxygen consumption rate along the length of the core and the relation between carbon age and oxidation rate. This approach had two advantages relative to calculating the reaction rate for oxygen only from profile curvature within discrete intervals (*13*). First, we avoided lumping high rates in the upper part of an interval with lower rates in the lower part, because this invariably leads to underestimation of the rates at the top of the interval. This is critical near the sediment surface, where the rates changed rapidly. Second, the relation between carbon reactivity and age can be used to predict how burial velocity influences the depth distribution of oxygen consumption and thus oxygen distribution (see below).

The deep oxygen profiles in the North Pacific Gyre were modeled well by assuming that the degradability of organic matter in the sediment follows a simple power law function of carbon

[1]Center for Geomicrobiology, Department of Bioscience, Aarhus University, Ny Munkegade 116, 8000 Aarhus C, Denmark. [2]Institute of Earth and Environmental Science, University of Potsdam, Germany. [3]Graduate School of Oceanography, University of Rhode Island, Narragansett, RI 02882, USA.

*To whom correspondence should be addressed. E-mail: hans.roy@biology.au.dk

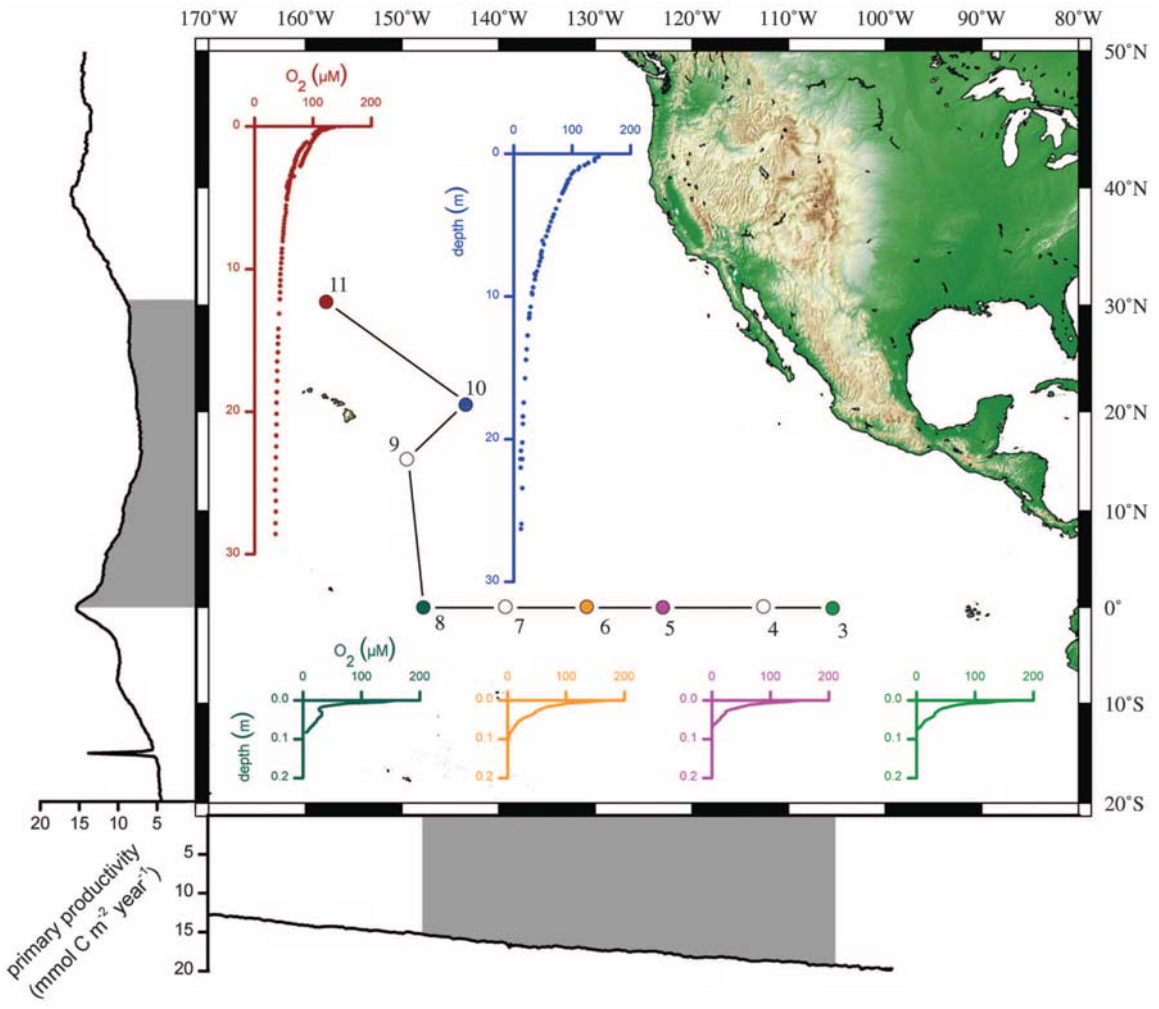


Fig. 1. Cruise track, sampling sites, and primary production along the cruise track. The primary production was estimated from Sea-WiFs remote-sensing data converted into integrated primary productivity averaged over 10 years by the Institute of Marine and Coastal Sciences Ocean Primary Productivity Team (Rutgers, State University of New Jersey) using the algorithms from (*25*). The shaded areas mark the data along the actual cruise track. The inserted graphs show the oxygen distribution in the sediment at the sampled sites. Sites 4 and 7 on the equator followed the trend in the other equatorial sites and are omitted for clarity only. At site 9 we retrieved only 4 m of core, but the profile was similar to sites 10 and 11. By site 11, we reoccupied the GPC-3 site that has been studied in great detail [e.g., (*14*)].

age (Fig. 2A). Oxygen penetration depth in the seabed beneath the North Pacific Gyre is more than 30 m. In the Atlantic gyres, with a similar organic carbon flux to the deep seafloor, oxygen penetrates only to 0.2 to 0.5 m in the sediment (*8*). The difference is a result of the low sedimentation rate in the North Pacific Gyre, where 90% of the organic matter mineralization in the 100-m deep oxic sediment at site 11 takes place in the top 6 cm. Carbon burial from the upper bioturbated zone into the deeper sediment is extremely slow, and the material is therefore highly refractory at depth. The fraction of total carbon mineralization that takes place below 1 m is about 1%.

The effect of low sedimentation rates on oxygen distribution was shown by using a numerical model similar to the one used to quantify oxygen consumption rates. Instead of feeding the measured carbon concentration profile into the model, we imposed a flux of carbon to the sediment surface. Burial was implemented as a downward advective term. Carbon and oxygen consumption were implemented by Eq. 1 with the fit parameters determined from site 11. We then varied the burial rate and calculated steady-state oxygen profiles (Fig. 3).

The shape of the theoretical oxygen distributions are close to those observed in the North Pacific Gyre (Fig. 3). All modeled profiles have the same oxygen gradient, that is, total oxygen flux, at the sediment-water interface because the scenarios are modeled with the same carbon input and because practically all the carbon is mineralized. At decreasing sedimentation rates, oxygen penetrates deeper because relatively less mineralization happens far below the surface. A shorter distance of diffusion between the sediment surface and the depth of mineralization caused less overall depletion of oxygen, although the total depth-integrated oxygen consumption was the same. A sudden shift in oxygen penetration, from dm range to the full depth of the sediment, occurs in a relatively narrow range of sedimentation rates of 1 to 5 mm per 1000 years. The sedimentation rate at site 11 has been stable at about 0.2 mm per 1000 years for the past 70 million years, with a moderate increase occurring during the most recent 2.5 million years (*14*).

The total sediment thickness at the core sites in the North Pacific Gyre is 88 to 100 m (*15*). Extrapolation of the oxygen gradient at the bottom of the cores suggests that the entire sediment column is oxic, at least at site 11. The fully oxic sediment column excludes anaerobic metabolic pathways that normally dominate subsurface mineralization of organic matter in the seabed. This provides a unique opportunity to quantify carbon mineralization rates as a function of sediment age across a long time span. Both traditional nonparametric modeling of oxygen consumption rates from profile curvature (*13*) and the applied power law model yielded volumetric oxygen consumption rates that decreased from 10 μmol O_2 liter^{-1} year^{-1} at a depth of 0.5 m below the sediment-water interface to 0.001 μmol O_2 liter^{-1}

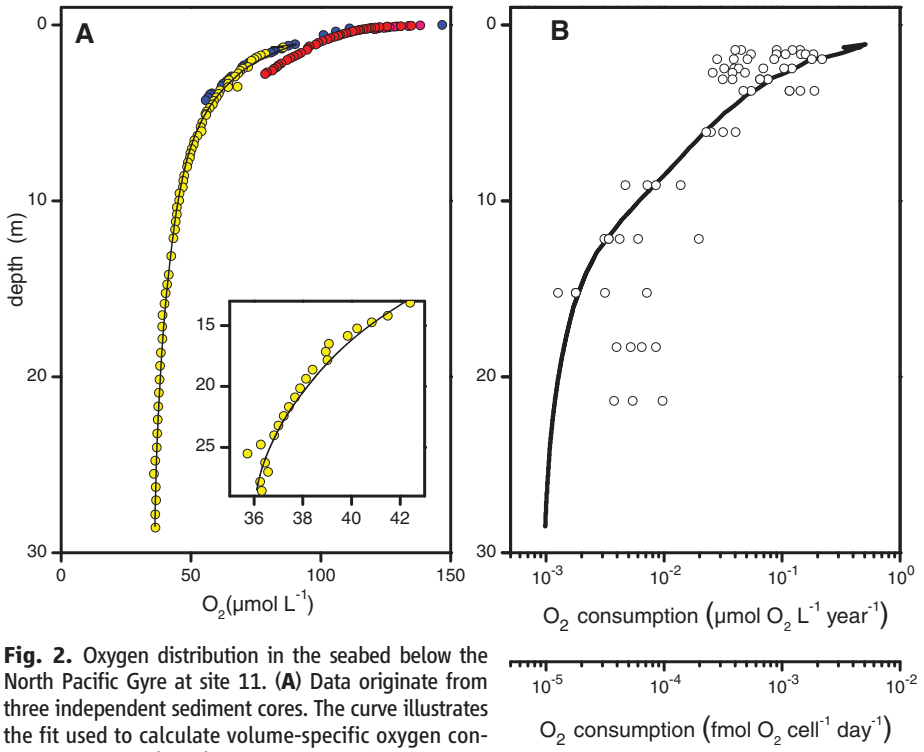


Fig. 2. Oxygen distribution in the seabed below the North Pacific Gyre at site 11. (**A**) Data originate from three independent sediment cores. The curve illustrates the fit used to calculate volume-specific oxygen consumption rates. (Inset) Data and fit for lower part of core. (**B**) Curve is the modeled volumetric oxygen consumption rates. Open circles are the cell-specific oxygen consumption rates obtained by dividing the volumetric oxygen consumption rate by cell counts.

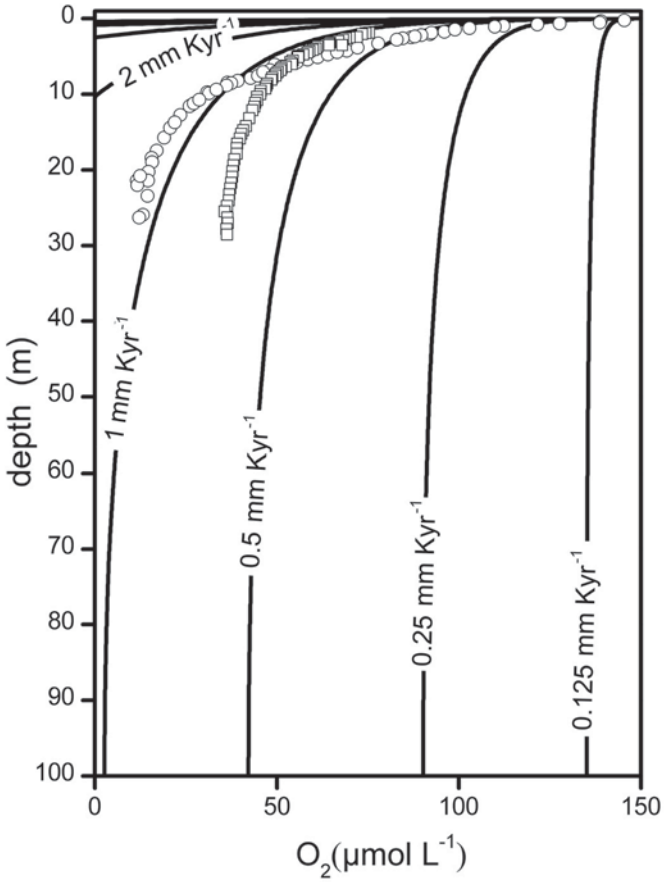


Fig. 3. Oxygen profiles modeled from a constant influx of organic material but varying sediment accumulation rates. All profiles represent the same oxygen uptake rate at the sediment surface. Symbols show measured oxygen profiles from site 9 (circles) and site 11 (squares).

year^{-1} in sediment older than 66 million years, that is, buried below 20 to 30 m at site 11 (*16*) (Fig. 2B).

The fit to a simple power law model (Fig. 2A) gave an exponent of –1.7, in contrast with an exponent of –1 reported by (*12*). Other continuously decreasing functions could have been used to describe the trend in the data, for example, a reactive continuum model that assumes exponential depletion of multiple hypothetical pools of complex organic matter (*17*, *18*). That approach is comparable to ours because a power law can be expressed as a sum of many exponential functions. We prefer the power law for its simplicity and because a model based on changing chemical properties of dead organic matter agreed well with the chemical alterations that occurred during aging and maturation.

The number of prokaryotic cells in the surface sediment of the North Pacific Gyre was 10^8 cells cm^{-3}. The cell counts decreased along the length of the core to 10^3 cells cm^{-3} at 20 m below the sea floor. Below that depth, the cell density was too low to enumerate by fluorescence microscope counting even after cell extraction (*19*). The cell density decreased relatively less down along the core than the volumetric oxygen consumption rate. The mean oxygen consumption rate per cell therefore decreased with increasing sediment depth and age (Fig. 2B). The per-cell respiration rate appeared to stabilize at around 10^{-3} fmol cell^{-1} day^{-1}. This overlaps with the range of cell-specific sulfate reduction rates in coastal subsurface sediments (*20*, *21*), but it is 3 orders of magnitude below the cell-specific respiration rate of anaerobic heterotrophs in pure culture (*22*) and below the respiration rate devoted to maintenance in the slowest-growing chemostat cultures (*23*). Thus, life in the subsurface is probably more similar to cultures in long-term stationary state (*24*) than to growing cultures. The higher cell-specific respiration rates reported previously from deeply oxygenated sediments in the South Pacific Gyre (*5*) are not comparable because those are average rates for the entire sediment column and are skewed upward by relatively high rates of metabolism at the sediment-water interface.

The similarity in mean metabolic rate per cell between sites with very different mineralization rates and different terminal electron acceptors suggests that these microbial communities may be living at the minimum energy flux needed for prokaryotic cells to subsist and that the total available energy flux ultimately controls the microbial community size in the deep biosphere.

References and Notes

1. R. J. Parkes *et al.*, *Nature* **371**, 410 (1994).
2. A. Schippers *et al.*, *Nature* **433**, 861 (2005).
3. B. B. Jørgensen, S. D'Hondt, *Science* **314**, 932 (2006).
4. J. P. Fischer, T. G. Ferdelman, S. D'Hondt, H. Røy, F. Wenzhofer, *Biogeosciences* **6**, 1467 (2009).
5. S. D'Hondt *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 11651 (2009).
6. W. J. Cai, F. L. Sayles, *Mar. Chem.* **52**, 123 (1996).
7. H. Rasmussen, B. B. Jørgensen, *Mar. Ecol. Prog. Ser.* **81**, 289 (1992).
8. F. Wenzhöfer, R. N. Glud, *Deep-Sea Res.* **49**, 1255 (2002).
9. F. Wenzhöfer, O. Holby, O. Kohls, *Deep-Sea Res.* **48**, 1741 (2001).
10. R. A. Jahnke, G. A. Jackson, *Nature* **329**, 621 (1987).
11. A. N. Antia *et al.*, *Global Biogeochem. Cycles* **15**, 845 (2001).
12. J. J. Middelburg, *Geochim. Cosmochim. Acta* **53**, 1577 (1989).
13. P. Berg, N. Risgaard-Petersen, S. Rysgaard, *Limnol. Oceanogr.* **43**, 1500 (1998).
14. F. T. Kyte, M. Leinen, G. Ross Heath, L. Zhou, *Geochim. Cosmochim. Acta* **57**, 1719 (1993).
15. D. L. Divins, Total Sediment Thickness of the World's Oceans and Marginal Seas, National Geophysical Data Center (NGDC), www.ngdc.noaa.gov/mgg/sedthick/sedthick.html (2009).
16. F. T. Kyte, J. T. Wasson, *Science* **232**, 1225 (1986).
17. A. Berner, *Colloq. Int. CNRS* **293**, 35 (1980).
18. B. P. Boudreau, B. R. Ruddick, *Am. J. Sci.* **291**, 507 (1991).
19. J. Kallmeyer, D. C. Smith, A. J. Spivack, S. D'Hondt, *Limnol. Oceanogr. Methods* **6**, 236 (2008).
20. K. Ravenschlag, K. Sahm, C. Knoblauch, B. B. Jørgensen, R. Amann, *Appl. Environ. Microbiol.* **66**, 3592 (2000).
21. L. Holmkvist, T. G. Ferdelman, B. B. Jørgensen, *Geochim. Cosmochim. Acta* **75**, 3581 (2011).
22. C. Knoblauch, B. B. Jørgensen, J. Harder, *Appl. Environ. Microbiol.* **65**, 4230 (1999).
23. P. van Bodegom, *Microb. Ecol.* **53**, 513 (2007).
24. S. E. Finkel, *Nat. Rev. Microbiol.* **4**, 113 (2006).
25. M. J. Behrenfeld, P. G. Falkowski, *Limnol. Oceanogr.* **42**, 1479 (1997).

Acknowledgments: The assistance from E. Caporelli, B. Costello, and E. Benway at Woods Hole Oceanographic Institution (WHOI) was crucial for our participation in the R/V *Knorr* cruise. We thank B. Gribsholt, the shipboard science party, the WHOI long-coring team, and the crew of the R/V *Knorr* for cooperation at sea. Net primary production data were provided by R. O'Malley from Oregon State University (www.science.oregonstate.edu/ocean.productivity/index.php). Our study was funded by the Danish National Research Foundation, the German Max Planck Society, Aarhus University, the German Federal Ministry of Research and Education via the GeoEnergie Project, and the U.S NSF (OCE grant 0752336 and the NSF-funded C-DEBI Science and Technology Center). This is C-DEBI publication number 130.

Supplementary Materials

www.sciencemag.org/cgi/content/full/336/6083/922/DC1
Materials and Methods
Table S1
References (*26*, *27*)

20 January 2012; accepted 17 April 2012
10.1126/science.1219424

Multiple Spectral Inputs Improve Motion Discrimination in the *Drosophila* Visual System

Trevor J. Wardill,[1]* Olivier List,[1]* Xiaofeng Li,[1,2]* Sidhartha Dongre,[1] Marie McCulloch,[1] Chun-Yuan Ting,[3] Cahir J. O'Kane,[4] Shiming Tang,[2] Chi-Hon Lee,[3] Roger C. Hardie,[5] Mikko Juusola[1,2]†

Color and motion information are thought to be channeled through separate neural pathways, but it remains unclear whether and how these pathways interact to improve motion perception. In insects, such as *Drosophila*, it has long been believed that motion information is fed exclusively by one spectral class of photoreceptor, so-called R1 to R6 cells; whereas R7 and R8 photoreceptors, which exist in multiple spectral classes, subserve color vision. Here, we report that R7 and R8 also contribute to the motion pathway. By using electrophysiological, optical, and behavioral assays, we found that R7/R8 information converge with and shape the motion pathway output, explaining flies' broadly tuned optomotor behavior by its composite responses. Our results demonstrate that inputs from photoreceptors of different spectral sensitivities improve motion discrimination, increasing robustness of perception.

The integration of color and motion signals at the neural circuit level is poorly understood, but combining these inputs is thought to improve perceptual discrimination (*1*, *2*), reducing uncertainty to make judgments and guide actions in the world (*3*, *4*). However, it remains elusive to what extent information from different spectral channels (*5*, *6*) is exploited to improve motion detection and orienting behavior (*7*–*10*). The modular architecture of the *Drosophila* visual system, with its genetic malleability and accessibility to in vivo recordings (*11*–*14*), makes it a powerful model to study such circuit computations. Its multifaceted compound eyes are composed of ~750 ommatidia, each containing photoreceptor cells with a suite of spectral sensitivities (*6*, *15*, *16*). Light from a point in space excites eight photoreceptors, R1 to R8, distributed across seven ommatidia (Fig. 1A) (*6*). Because each of these photoreceptors innervate either the parallel motion or color pathway, they are thought to subserve specific visual behaviors (*17*, *18*), and for over 30 years most researchers have considered insect motion vision to be mediated exclusively by the single spectral class, R1-R6 (*9*, *11*, *17*–*20*).

[1]Department of Biomedical Science, University of Sheffield, Sheffield S10 2TN, UK. [2]State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100875, China. [3]Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892–5431, USA. [4]Department of Genetics, University of Cambridge, Cambridge CB2 3EH, UK. [5]Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge CB3 2EG, UK.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: m.juusola@sheffield.ac.uk

Outer photoreceptors R1-R6 (*6*), expressing blue-green Rh1-opsin (*21*) and ultraviolet (UV)-sensitizing pigment (*22*), mediate motion detection (*20*) by innervating (*23*) large monopolar cells (LMCs, L1 and L2) (*11*, *18*) in the first optic neuropile, the lamina. These neurons, which thus should have the same spectral sensitivity as R1-R6, project to the second optic neuropile, the medulla (*24*). The signals transmit from medulla output layers to the motion-sensitive tangential cells of the lobula plate (LPTCs), feeding the fly's optomotor behavior to changing optic flow fields (*25*). Light also excites central photoreceptors, R7 and underneath it R8, initiating the color pathway. R7s express UV-sensitive opsins (R7p:Rh3; R7y:Rh4), whereas R8s express blue- (R8p:Rh5) or green-yellow–sensitive opsins (R8y:Rh6), having 30:70 retinal distributions, respectively (*6*, *16*). R7s/R8s innervate separate medulla layers (*24*), where some contribution to their processing comes from R1-R6 via monopolar neurons L3 (*17*).

Although motion information appears routed and processed through dedicated on and off channels (*11*, *26*, *27*), in principle R1-R6 channels might cross-talk with the color channels in places. Ultrastructural studies imply interactions in the medulla, with synapses between R7/R8 photoreceptors and LMCs (*24*), but possibly also in the lamina, where gap junctions (*28*, *29*) and synapses (*13*) form sophisticated local processing networks (*30*). Here, combining information from different channels could improve the sensitivity, precision, and robustness of their neural messages. However, such interactions are hard to investigate because the spectral sensitivities of the color and motion channels overlap extensively. This interference makes it difficult to activate one pathway without activating the other.

To overcome this problem, we generated flies, referred to as UV flies, in which the R1-R6 blue-green opsin (Rh1, also known as *ninaE*) was mutated [$ninaE^{8}$, also $ninaE^{P334}$ (*31*)], making it functionally inactive, and a UV opsin (Rh3) (*32*) was expressed in R1-R6 photoreceptors in its place (Fig. 1B). Mutant $ninaE^{8}$ flies were selected because they have intact R1-R6 ultrastructure (*33*) while their R7/R8 photoreceptors function normally (Fig. 1F and fig. S2); by contrast, many alternative *ninaE* lines carry additional mutations (fig. S2 and table S1), which may blind them (*34*). Hence, by using a monochromator light source, R8s of UV flies could now be independently excited by long-wavelength light (≥480 nm), whereas R1-R6s should only respond to short-wavelength light (<480 nm).

Whole-cell recordings of dissociated R1-R6 photoreceptors from these UV flies indicated that, apart from the expected UV sensitivity of Rh3-opsin (Fig. 1C and fig. S3), their phototransduction dynamics (Fig. 1D) approximate those of the wild-type photoreceptors (*35*). The intact retina ultrastructure (Fig. 1E and fig. S1B), in vivo electroretinograms (Fig. 1F), and response dynamics (fig. S4) indicated undistorted signal transmission from R1-R6 to LMCs, suggesting that the underlying circuit computations were normal. Therefore, by using green-amber stimulation, which is essentially invisible to these UV-sensitive R1-R6s but excites R8 photoreceptors, we could examine whether there is cross-talk between different photoreceptor classes.

Intracellular recordings (Fig. 2A) in R1-R6s and LMCs from UV flies showed that input from R8 photoreceptors can already be detected in the lamina circuits. In addition to their predicted UV sensitivity (Fig. 2B), ~30% of R1-R6 photoreceptors (12/36, putatively axons) and all LMCs (28/28) responded to longer wavelength flashes, either by brief depolarizations (LMCs 7%; Fig. 2C, left and right) or hyperpolarizations (LMCs 93%; Fig. 2C, middle), mapping R8y or R8p photoreceptors' spectral range (Fig. 2D; corresponding nomograms). These relatively small

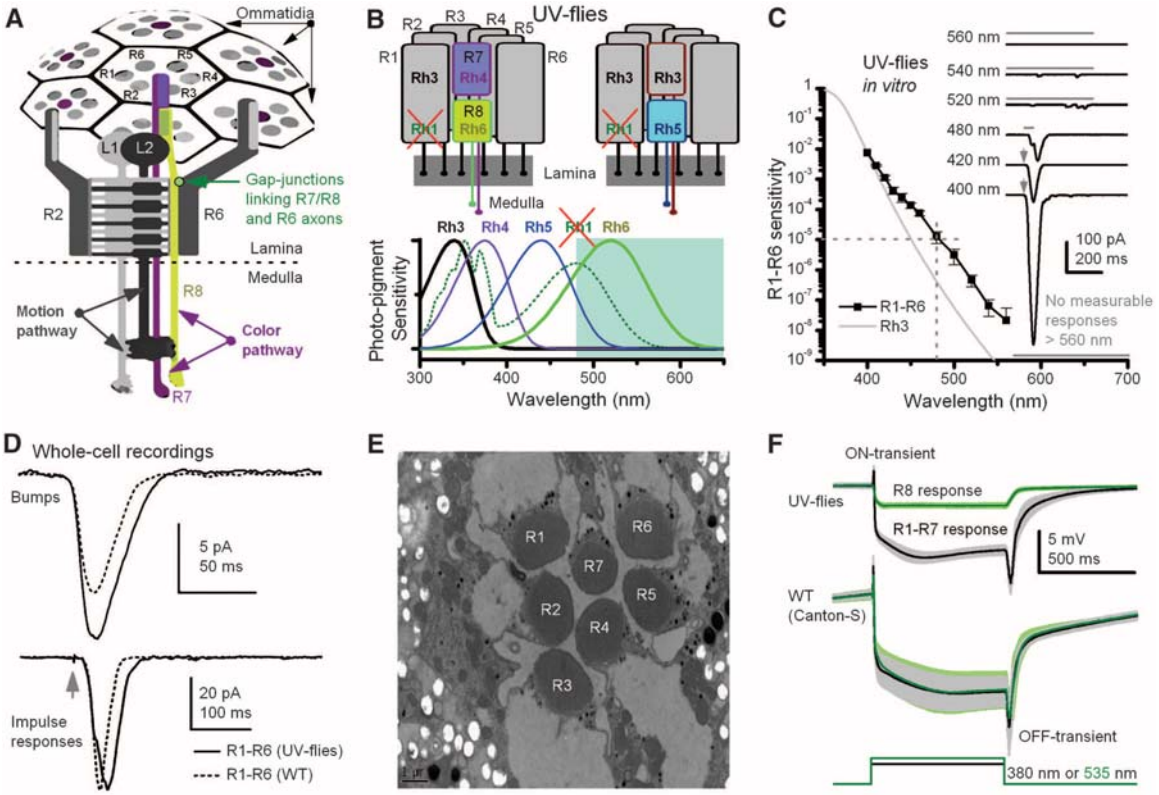


Fig. 1. Manipulating spectral sensitivity of the motion pathway to elucidate circuit computations. (**A**) Schematic of fly photoreceptors innervating motion and color pathways. A light point is sampled by six outer photoreceptors (R1-R6) in neighboring ommatidia and by central R7/R8 photoreceptors. These R1-R6s innervate LMCs (L1 and L2) in the first optic neuropile, lamina, whereas R7/R8 synapse in the second neuropile, medulla. Gap junctions link R7/R8 to R6 just before R1-R6 are brought together (superpositioned) for synaptic transmission (*28*). (**B**) In UV flies, UV-sensitive Rh3-opsin is expressed in R1-R6s, which contain nonfunctional Rh1-opsin ($ninaE^{8}$). This circumvents the spectral overlap between motion and R8 pathways. (**C**) Spectral sensitivity of light-induced currents (inset) in dissociated R1-R6s of UV flies, diminishing <1/100,000 at ≥480 nm (mean ± SD; four photoreceptors). (Inset) Example responses to 2-ms flashes (arrows) or 50- or 500-ms stimuli (bars). (**D**) Although single-photon responses (bumps) can be larger than those of WT (Canton-S) and impulse responses last slightly longer, their responses are very similar to their WT counterparts, indicating normal-like phototransduction (fig. S4, A and B) (*35*). (**E**) Photoreceptors of UV flies have normal-like ultrastructure. (**F**) Electroretinograms (ERGs) show comparable dynamics to those of WT flies: Preferred colors evoke large receptor components and on and off transients, indicating normal synaptic transmission from R1-R6 to LMCs (means ± SD; six flies). With UV flies, one can separate the responses for green (R8s) and UV (R1-R7s).

and slightly delayed responses implied that their input probably came through gap junctions or synapses at the level of R1-R6 terminals (fig. S5A) and, in somatic photoreceptor recordings, reflected signal degradation by back propagation. Their opposing polarities confirmed that these signals were neither recording artifacts nor field potentials (fig. S6). Thus, the spectral sensitivity of many R1-R6s and every tested LMC was selectively boosted in the 460- to 600-nm range (Fig. 2D).

These findings support a wiring model (Fig. 2E) in which one photoreceptor (R6 or neighboring R1) receives information from either a R8y or R8p cell, in keeping with results from larger dipteran flies (fig. S7) (*28*, *29*). This input would then spread decrementally between R1-R6 via gap junctions (*29*, *36*), showing diminishing green-yellow sensitivity the farther a recorded photoreceptor is from a R6 in the cartridge (fig. S6A). Because every R1-R6 forms synapses with every L1-L3 (*30*), all LMCs in the same cartridge should receive this information. Correspondingly, all LMCs showed larger responses to green-yellow than those recorded from R1-R6s (Fig. 2D). Nonetheless, LMCs' responsiveness to long wavelengths (460 to 620 nm) was ~10 times less (9.5% ± 7.0, SD; n = 28; range 2.8 to 34.4%) than that to UV light (100%; 300 to 440 nm), consistent with the model that they receive green-yellow inputs mainly from one R6 instead of the normal R1-R6. In some LMC recordings, the speed and size of depolarizations to green-yellow stimuli (Fig. 2, C and E, middle) implied that the R8 axon forms gap junctions with one LMC and that such signals feed back to selected R1-R6s (fig. S5A), inhibiting their output.

To assess the contribution of R7/R8 input to neural signaling between R1-R6 and LMCs, we crossed our UV flies into a phototransduction mutant background (*norpA*^{36}), which rendered all photoreceptors blind, and then rescued only R1-R6 by using a wild-type *norpA* transgene, resulting in structurally intact but light-insensitive R7/R8 photoreceptors (Fig. 2F). Because this manipulation prevented R7/R8 phototransduction, we could quantify information transfer from the color pathway to R1-R6 and LMCs by comparing their outputs with and without R7/R8 inputs, but with effectively wild-type R1-R6 dynamics and impedances (fig. S4, C and D). Without R7/R8 inputs, R1-R6s and LMCs could respond only to 300- to 440-nm stimuli (Fig. 2, G to I), following the spectral sensitivity of the UV-opsin expressed in R1-R6 photoreceptors (Fig. 1C). When R7/R8 cells were not participating in the circuit, the responses of R1-R6s were briefer when saturated with a UV impulse (340 ± 20 nm), which maximally excites R1-R6s and R7s but only slightly R8s (Fig. 3A and fig. S5B). The briefer responses of R1-R6s [half-width: P = 0.009, one-way analysis of variance (ANOVA)] and LMCs (P = 0.021) as seen over the tested spectral range in recordings probably reflected the lack of R7/R8 input in their processing.

We next quantified how R7/R8 inputs affect encoding of R1-R6s (Fig. 3B) by using repeated stimulation with naturalistic intensity patterns in preparations where R7/R8s either functioned normally or were light-insensitive. With functionally intact R7/R8s, the signal-to-noise ratio of R1-R6 photoreceptors for low-frequency UV stimulation

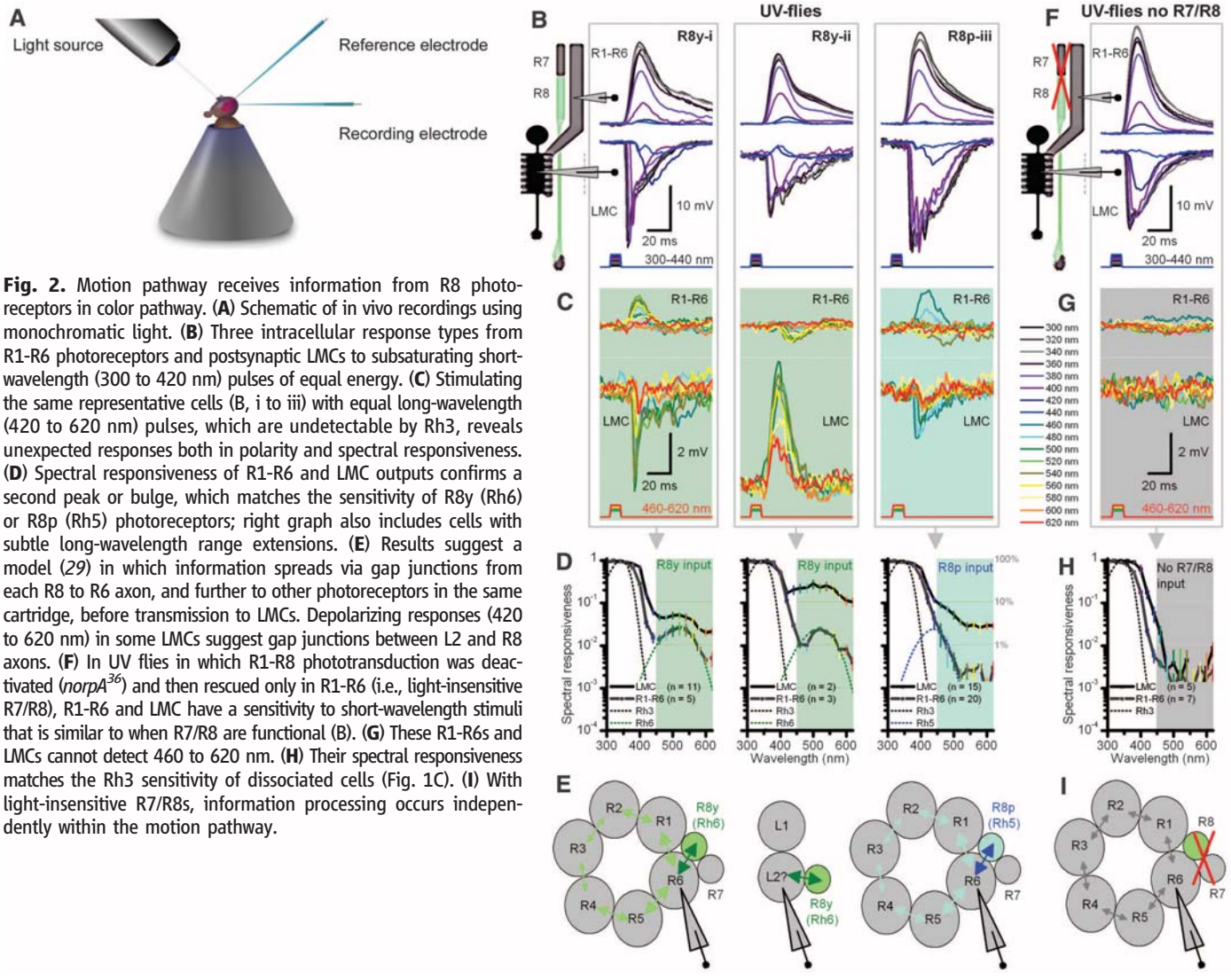


Fig. 2. Motion pathway receives information from R8 photoreceptors in color pathway. (**A**) Schematic of in vivo recordings using monochromatic light. (**B**) Three intracellular response types from R1-R6 photoreceptors and postsynaptic LMCs to subsaturating short-wavelength (300 to 420 nm) pulses of equal energy. (**C**) Stimulating the same representative cells (B, i to iii) with equal long-wavelength (420 to 620 nm) pulses, which are undetectable by Rh3, reveals unexpected responses both in polarity and spectral responsiveness. (**D**) Spectral responsiveness of R1-R6 and LMC outputs confirms a second peak or bulge, which matches the sensitivity of R8y (Rh6) or R8p (Rh5) photoreceptors; right graph also includes cells with subtle long-wavelength range extensions. (**E**) Results suggest a model (*29*) in which information spreads via gap junctions from each R8 to R6 axon, and further to other photoreceptors in the same cartridge, before transmission to LMCs. Depolarizing responses (420 to 620 nm) in some LMCs suggest gap junctions between L2 and R8 axons. (**F**) In UV flies in which R1-R8 phototransduction was deactivated (*norpA*^{36}) and then rescued only in R1-R6 (i.e., light-insensitive R7/R8), R1-R6 and LMC have a sensitivity to short-wavelength stimuli that is similar to when R7/R8 are functional (B). (**G**) These R1-R6s and LMCs cannot detect 460 to 620 nm. (**H**) Their spectral responsiveness matches the Rh3 sensitivity of dissociated cells (Fig. 1C). (**I**) With light-insensitive R7/R8s, information processing occurs independently within the motion pathway.

(<10 Hz, 390 nm) was increased by ~100% (Fig. 3C; $P < 0.05$, one-way ANOVA); on average, this equals an extra 10 to 20 bit/s (or ~5 to 10%; inset) on a R1-R6's information transfer (*37*). However, transfer of long-wavelength (595 nm) information in presumptive R6s (Fig. 3, D and E)—that is, ~12% of R1-R6 photoreceptors, which receive the direct feed from R8y photoreceptors (*28*, *29*) of the color pathway [70% lamina cartridges (*16*)]—could be as high as 70% compared with that of a similar UV stimulus in the same cells (Fig. 3F), indicating efficient communication between R8 and R6 axons. Thus, R7/R8 inputs boost the amplitude range of relatively slow voltage deflections to capture better the temporal structure of the dominant slow contrast changes in *Drosophila*'s natural environment (*38*).

We then used *ninaE*^{8} flies to determine whether R7/R8 input alone is sufficient to drive LMC output. In vitro patch-clamp experiments showed no macroscopic light responses in their R1-R6s (Fig. 3G) but essentially normal K^+ conductances (fig. S4F) and only moderately reduced capacitance. In vivo, we struggled to record light-induced voltage responses in the retina from R1-R6 somata. However, their axons closer to (or within) the lamina responded to a broad spectral range of light pulses with small rapidly rising depolarizations (Fig. 3H and fig. S5C), implying that they receive information from R7/R8s (*28*, *29*) through functional contacts, as in UV flies (Fig. 2C and fig. S5A). This information was faithfully transmitted to LMCs, all of which responded to light. Notably, the responses of most R1-R6s (9/11) and all postsynaptic LMCs (9/9) (Fig. 3H) followed the distinctive spectral sensitivity of R7p/R8p or R7y/R8y pairs (Fig. 3I) (*6*, *15*, *16*); two R1-R6s were UV-sensitive (fig. S5D). This suggests that presynaptic coupling between R1-R6 and R7/R8 axons can extend neural superposition in most lamina cartridges to eight photoreceptors (instead of the assumed six), further improving information transfer to LMCs (*13*). We conclude that R7 and R8 inputs alone in *ninaE*^{8} and UV flies can drive the synaptic output of R1-R6 terminals (Fig. 3J) to activate the on and off channels [L1 and L2 monopolar cells, respectively (*11*)] for motion detection.

Although these results (Figs. 2 and 3) demonstrate the spectral and temporal push-and-pull of R7/R8s input on both R1-R6 and LMC outputs, supporting a wiring model in which color pathways refine signaling in on and off channels, it remained unclear whether this information is sufficient or necessary for motion perception. To resolve these questions, we next examined *Drosophila*'s optomotor behavior.

By using a classic flight simulator system (*18*, *20*), we quantified how altering the spectral sensitivity or availability of photoreceptor inputs affect the flies' behavioral responses to colored field rotations. Figure 4 shows the average optomotor responses (yaw torque) of tethered flying wild-type (WT) flies, UV flies, and specific mutants

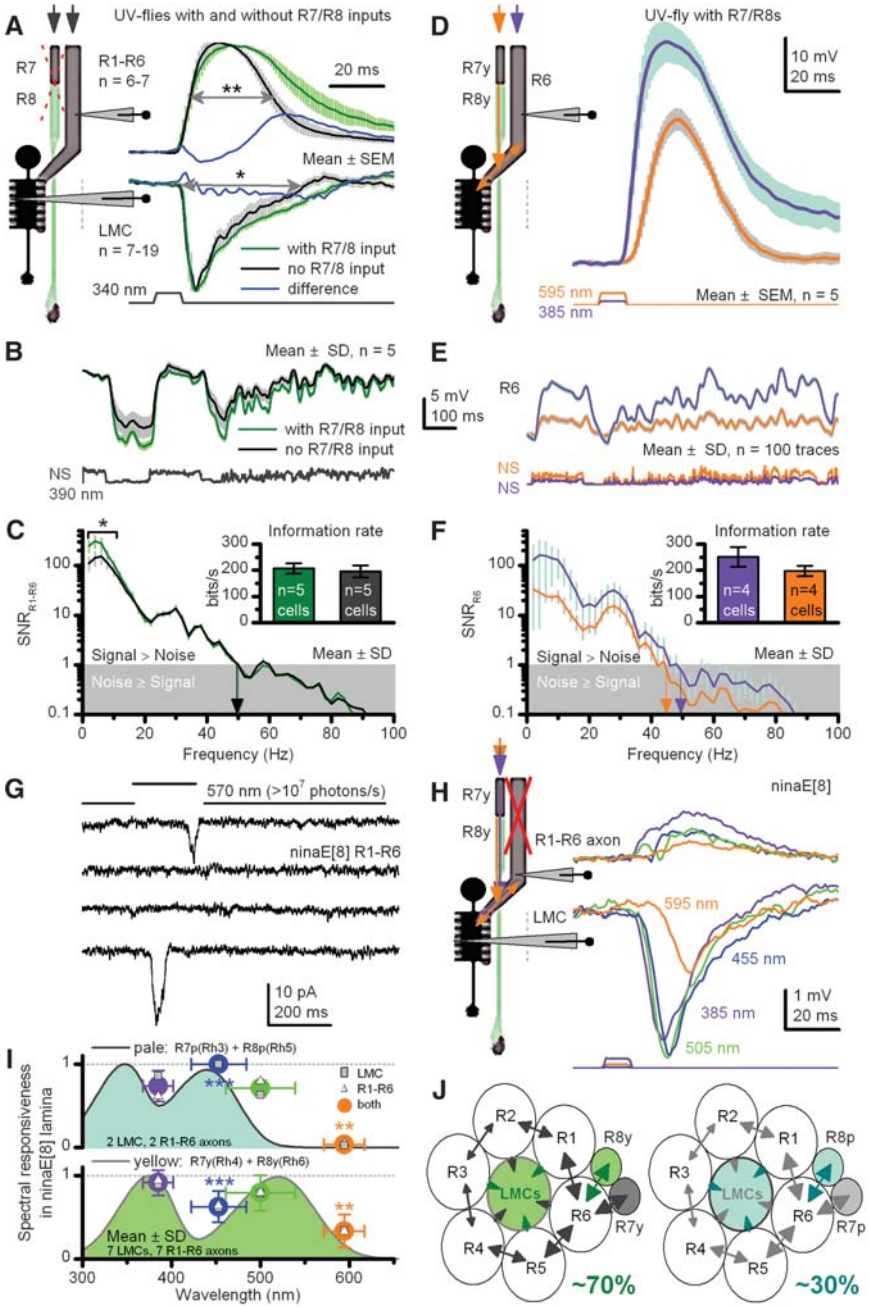


Fig. 3. R7/R8 photoreceptors transmit information to a motion pathway shaping R1-R6 and LMC outputs. (**A**) Normalized R1-R6 and LMC outputs to a saturating UV impulse in UV flies with or without light-sensitive R7/R8s. Wider responses suggest extra R7 inputs. Differences highlight the change over time between the genotypes. (**B**) R1-R6 outputs to naturalistic UV-intensity series; NS, with or without light-sensitive R7/R8s. (**C**) Corresponding signal-to-noise ratios (SNRs) and information transfers (inset). (**D**) Apart from the three response types in Fig. 2C, ~10% of R1-R6s from UV flies also responded strongly (>25 mV) to amber, signifying direct R8y input to R6 (*29*) (figs. S6 and S7). (**E**) Responses to UV and amber NS [different from (B)] in one presumptive R6. (**F**) High SNR with amber responses, containing ~70% of information of maximal UV responses (inset, three R6s). (**G**) Whole-cell recordings reveal that *ninaE*^{8} R1-R6s are profoundly light-insensitive (fig. S3B), lacking macroscopic light-induced current; ultrabright green-yellow light rarely evokes single-photon responses (two shown). (**H**) An in vivo example where presynaptic *ninaE*^{8} R1-R6 axon and postsynaptic LMC generate small responses, following R7/R8s' sensitivity, here R7y/R8y; similar recordings from nine LMCs and nine R1-R6s (fig. S5C). (**I**) Responsiveness of *ninaE*^{8} LMCs and R1-R6 axons track R7/R8-nomogram pairs. (**J**) Input to each lamina cartridge comes from R7, R8, R7y/R8y, or R7p/R8p pairs through gap junctions to R6 axons to drive synaptic output to LMCs. *$P < 0.05$; **$P < 0.01$, ***$P < 0.005$, one-way ANOVA.

to black-and-white, black-and-blue, or black-and-amber field rotations, which were chosen to evaluate the relative contribution of R7/R8 photoreceptors to their motion detection (control and individual responses are in figs. S9 and S10).

If all R7s/R8s supplied inputs to the motion pathway, then UV flies should show spectrally broad optomotor behavior. Indeed, their responses to the colored motion stimuli covered the range of WT responses (Fig. 4A), including amber motion, which in UV flies is only visible to their R8y photoreceptors. Naturally, WT R1-R6s could resolve amber motion better, because of their broadly tuned Rh1-opsin (Fig. 1B) and further boosting by R8y input via functional contacts, before transmission to the LMC on and off channels (*11*, *26*). Nonetheless, the overall similar response dynamics of UV and WT flies strongly suggest that the underlying computations in their motion-detection and flight-control circuits are similar (Fig. 4A, left).

Because input from R7/R8s alone can drive the synaptic terminals of otherwise light-insensitive R1-R6s (Fig. 3, H and I), our model predicts that *ninaE*^{8} flies, which only have light-sensitive R7/R8 photoreceptors, should also respond to field rotations. Indeed, *ninaE*^{8} flies also showed robust optomotor responses (Fig. 4B), consistent with efficient information transfer from R7/R8s to R6 cells (Fig. 3F). As predicted, the optomotor response was weakest with amber illumination when only single photoreceptor type (R8y), with only a small fraction of its spectral range being stimulated, drove on and off channels (amber motion), in agreement with intracellular recordings (Fig. 3I and fig. S5C).

We then asked whether R7/R8 inputs are necessary for motion detection. Flies in which R1-R6 photoreceptors expressed either Rh1-opsin (wild-type) or Rh3-opsin (UV) but had light-insensitive color (R7/R8) pathways attempted to follow visual motion (Fig. 4C). However, these torque responses were weaker than those of WT (P = 0.012) and UV flies (P = 0.007; also faster, P = 0.0016, two-way t test) (Fig. 4A), consistent with the briefer voltage responses of their R1-R6 and LMCs (Fig. 3A). Thus, R1-R6s inputs into motion pathway alone are sufficient for motion detection (*18*, *20*, *39*, *40*), but R7 and R8 inputs boost the slow response components, adjusting the size, speed, and spectral range of optomotor behavior.

Next, to dissect the contribution of the different color channels in optomotor behavior, we

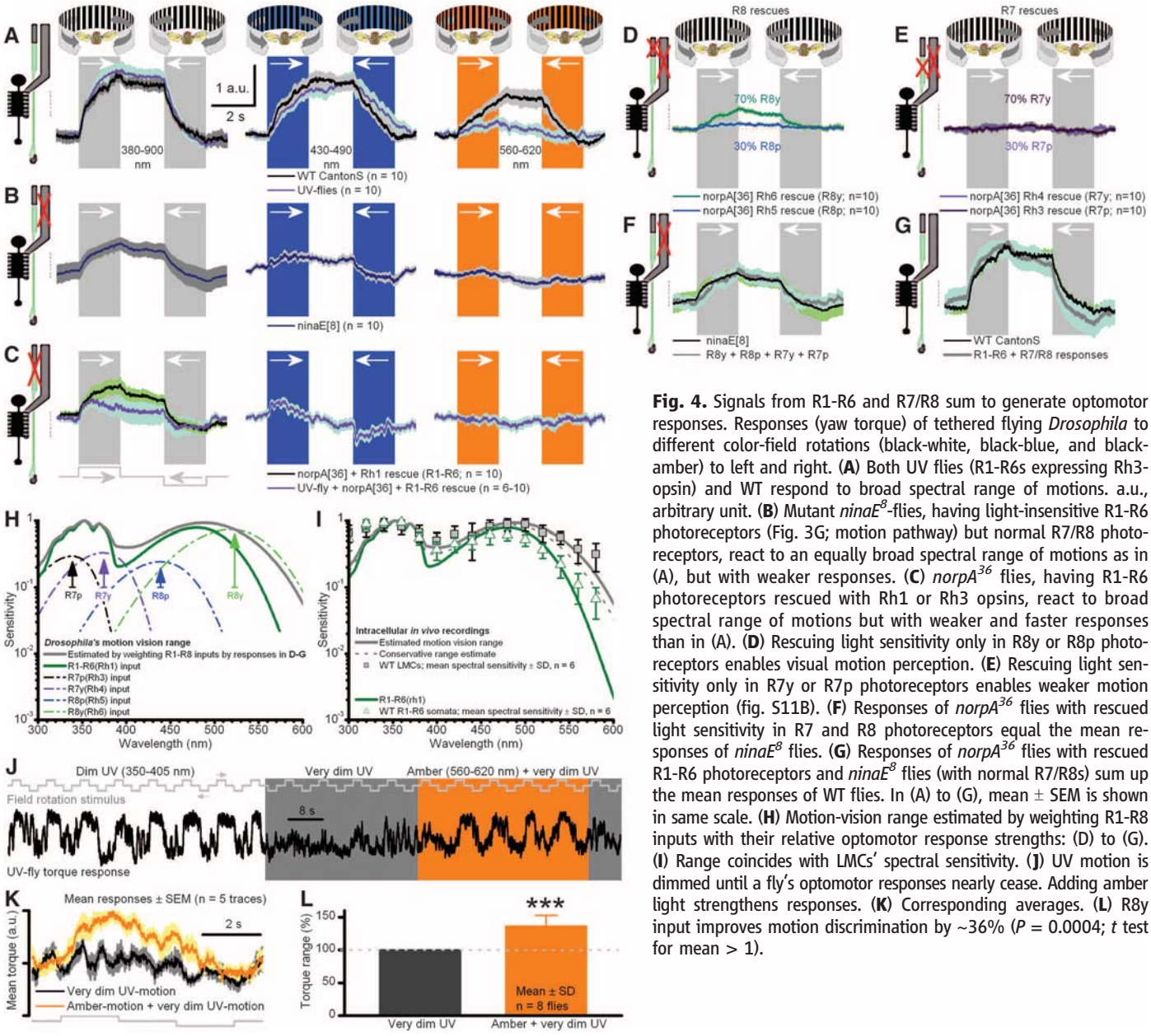


Fig. 4. Signals from R1-R6 and R7/R8 sum to generate optomotor responses. Responses (yaw torque) of tethered flying *Drosophila* to different color-field rotations (black-white, black-blue, and black-amber) to left and right. (**A**) Both UV flies (R1-R6s expressing Rh3-opsin) and WT respond to broad spectral range of motions. a.u., arbitrary unit. (**B**) Mutant *ninaE*^{8}-flies, having light-insensitive R1-R6 photoreceptors (Fig. 3G; motion pathway) but normal R7/R8 photoreceptors, react to an equally broad spectral range of motions as in (A), but with weaker responses. (**C**) *norpA*^{36} flies, having R1-R6 photoreceptors rescued with Rh1 or Rh3 opsins, react to broad spectral range of motions but with weaker and faster responses than in (A). (**D**) Rescuing light sensitivity only in R8y or R8p photoreceptors enables visual motion perception. (**E**) Rescuing light sensitivity only in R7y or R7p photoreceptors enables weaker motion perception (fig. S11B). (**F**) Responses of *norpA*^{36} flies with rescued light sensitivity in R7 and R8 photoreceptors equal the mean responses of *ninaE*^{8} flies. (**G**) Responses of *norpA*^{36} flies with rescued R1-R6 photoreceptors and *ninaE*^{8} flies (with normal R7/R8s) sum up the mean responses of WT flies. In (A) to (G), mean ± SEM is shown in same scale. (**H**) Motion-vision range estimated by weighting R1-R8 inputs with their relative optomotor response strengths: (D) to (G). (**I**) Range coincides with LMCs' spectral sensitivity. (**J**) UV motion is dimmed until a fly's optomotor responses nearly cease. Adding amber light strengthens responses. (**K**) Corresponding averages. (**L**) R8y input improves motion discrimination by ~36% (P = 0.0004; t test for mean > 1).

rescued R8y, R8p, R7y, or R7p photoreceptors individually with a *norpA* WT transgene under control of their corresponding opsin promoters in blind *norpA* mutants (fig. S8, A to D). After verifying through patch-clamp recordings that R1-R6 photoreceptors of these flies were light-insensitive (fig. S8, E and F), we tested their optomotor behavior to black-and-white (spectrally broad) field rotations (Fig. 4, D and E). The strength of their optomotor responses reflected broadly the retinal distribution of opsins (*6*, *15*, *16*), supporting our hypothesis that motion perception emerges from spectral integration of receptor outputs. When R8y(Rh6)s, which reside in most ommatidia, were rescued, these populations alone provided a fly with robust motion perception (normalized mean torque response of 100%), whereas the responses of flies, with other R7/R8 subtypes rescued, were smaller ($P < 0.05$, $n = 10$ flies, two-way t test; fig. S11): for R7y(Rh4), 44.8%; R8p(Rh5), 32.5%; and R7p(Rh3), 40.2%. These differences may reflect variable amounts, distributions, or strengths of functional contacts, which different R7 and R8 types make with the motion pathway in the lamina (*28*) and medulla circuits (*17*, *24*). The sum of optomotor responses from these four opsin-rescued genotypes matched the mean response of $ninaE^8$ flies (Fig. 4F), in which R7/R8 pairs are light-sensitive. Equally, the UV and $ninaE^8$ flies responses to amber motion, although weak (Fig. 4, A and B right), were similar, highlighting the robustness of R7/R8 input for motion detection. Lastly, the sum of responses from flies with only light-sensitive R7/R8s (Fig. 4B) or only light-sensitive R1-R6s (Fig. 4C) approximated the mean response of WT flies (Fig. 4G). This unexpectedly linear summation allowed us to estimate the spectral range of motion vision by scaling the spectral sensitivity (nomogram) of each photoreceptor class by its apparent contribution to WT optomotor responses (Fig. 4H). The estimated range approximates the spectral sensitivity of WT LMCs (Fig. 4I), whereas the sensitivity of WT R1-R6s is more like their Rh1 input, suggesting that, from the first synapse onward, combined inputs from photoreceptors of different spectral classes mediate *Drosophila*'s motion perception.

As a decisive test that R7/R8 inputs augment motion vision (Fig. 4J), we dimmed UV-lit field rotations to a level where UV flies' optomotor behavior nearly ceased and then shone additional amber light (595 nm), which is only detected by R8y photoreceptors, on the rotating field. Amber motion information strengthened the flies' torque responses, confirming that inputs from photoreceptors of different spectral sensitivities improve motion discrimination (Fig. 4, K and L).

Lastly, by using two-photon infrared microscopy to avoid stimulation of the photoreceptors, we imaged neural activity patterns in the motion-sensitive neurons (LPTCs) in the lobula plate of the *Drosophila* brain (figs. S12 to S14). The LPTCs of UV and WT flies were engineered to express genetically targeted calcium indicators (GCaMP3) that responded to changes in neural activity (fig. S12B). When such a fly viewed wide-field top-down and bottom-up movements of different colors, we found that individual LPTCs of both UV and WT flies preferred different motion directions (figs. S12, C and D; to S14). They responded to UV and amber-green image motions in comparable fashion; the responses to UV motion were stronger than those to amber motion [UV: DF/F (change in fluorescence/fluorescence) = 70.5% ± 19.3; amber: 33.7% ± 21.2, ± SD, $P = 0.001$, one-way ANOVA; $n = 10$ LPTCs from five UV flies].

We have shown that inputs from R7 and R8 photoreceptors converge with and shape R1-R6/LMC outputs, improving motion discrimination in *Drosophila*. These findings refute the long-standing hypothesis that motion detection in flies is mediated exclusively by one spectral class of photoreceptors (R1-R6) but agree with the circuit model, in which R7 and R8 photoreceptors feed information to R1-R6 terminals through gap-junctions (*28*, *29*). Besides extending and equalizing the spectral range of motion detection so that each spectral band has more similar probability of being used, R7/R8 inputs adjust the sensitivity (size and speed) of optomotor responses. Sensitivity increases as each on and off channel receives inputs from eight superpositioned photoreceptors, instead of six as previously reported, with R7/R8 inputs adding extra low-frequency information about slow- or low-contrast changes. Nonetheless, R7/R8 inputs improve robustness of motion vision; by piggy-backing to the motion pathway just before the first visual synapse, inputs from even a single R7/R8 receptor class enable motion perception.

These results also suggest an answer to the old open question: Why are R7y/R8y and R7p/R8p pairs randomly distributed across the retina? Although a color-blind movement detector is in many senses optimal (*41*), it cannot respond to movement of two-color patterns if the intensities of the colors provide no contrast to the input channels: the isoluminant point (*42*). Furthermore, in the spatial domain, the receptive fields of on and off channels of an optimal motion detector should be maximally correlated, that is, identical, for the two channels to filter the scene in the same way (*41*). In a previous study (*9*), the capacity of many WT flies to respond to opposing movements of blue-green color contrast (of different intensity ratios) never fell to zero; thus, such flies showed no absolute point of perceptual isoluminance, consistent with our findings here. This suggests that *Drosophila*'s motion-detection circuits may have evolved to minimize isoluminance effects by deriving its inputs from randomized combinations of spectral receptors, both within the same point in space and between neighboring points, to supply the on and off channels used for motion detection (*6*, *15*, *16*). Thus, these circuits provide simultaneous and robust encoding of multiple visual attributes from the ever-changing external world with limited wiring costs.

References and Notes

1. S. Nishida, J. Watanabe, I. Kuriki, T. Tokimoto, *Curr. Biol.* **17**, 366 (2007).
2. T. Takeuchi, K. K. De Valois, J. L. Hardy, *Vision Res.* **43**, 1159 (2003).
3. H. B. Barlow, *Behav. Brain Sci.* **24**, 602 (2001).
4. D. C. Knill, A. Pouget, *Trends Neurosci.* **27**, 712 (2004).
5. P. H. Schiller, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 17087 (2010).
6. R. C. Hardie, *J. Comp. Physiol. A* **129**, 19 (1979).
7. S. J. Cropper, S. M. Wuerger, *Behav. Cogn. Neurosci. Rev.* **4**, 192 (2005).
8. K. R. Gegenfurtner, M. J. Hawken, *Trends Neurosci.* **19**, 394 (1996).
9. S. Yamaguchi, R. Wolf, C. Desplan, M. Heisenberg, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 4910 (2008).
10. A. M. Derrington, *Curr. Biol.* **10**, R268 (2000).
11. M. Joesch, B. Schnell, S. V. Raghu, D. F. Reiff, A. Borst, *Nature* **468**, 300 (2010).
12. M. Juusola, R. C. Hardie, *J. Gen. Physiol.* **117**, 3 (2001).
13. L. Zheng *et al.*, *J. Gen. Physiol.* **127**, 495 (2006).
14. M. E. Chiappe, J. D. Seelig, M. B. Reiser, V. Jayaraman, *Curr. Biol.* **20**, 1470 (2010).
15. S. G. Britt, R. Feiler, K. Kirschfeld, C. S. Zuker, *Neuron* **11**, 29 (1993).
16. W. H. Chou *et al.*, *Neuron* **17**, 1101 (1996).
17. S. Gao *et al.*, *Neuron* **60**, 328 (2008).
18. J. Rister *et al.*, *Neuron* **56**, 155 (2007).
19. W. Kaiser, E. Liske, *J. Comp. Physiol.* **89**, 391 (1974).
20. M. Heisenberg, E. Buchner, *J. Comp. Physiol.* **117**, 127 (1977).
21. J. E. O'Tousa *et al.*, *Cell* **40**, 839 (1985).
22. K. Kirschfeld, N. Franceschini, B. Minke, *Nature* **269**, 386 (1977).
23. R. C. Hardie, *Nature* **339**, 704 (1989).
24. S. Y. Takemura, Z. Lu, I. A. Meinertzhagen, *J. Comp. Neurol.* **509**, 493 (2008).
25. B. Bausenwein, R. Wolf, M. Heisenberg, *J. Neurogenet.* **3**, 87 (1986).
26. G. G. de Polavieja, *Neural Comput.* **18**, 2102 (2006).
27. A. Riehle, N. Franceschini, *Exp. Brain Res.* **54**, 390 (1984).
28. S. R. Shaw, A. Fröhlich, I. A. Meinertzhagen, *Cell Tissue Res.* **257**, 295 (1989).
29. S. R. Shaw, *J. Exp. Biol.* **112**, 225 (1984).
30. I. A. Meinertzhagen, S. D. O'Neil, *J. Comp. Neurol.* **305**, 232 (1991).
31. N. J. Scavarda, J. O'tousa, W. L. Pak, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 4441 (1983).
32. R. Feiler *et al.*, *J. Neurosci.* **12**, 3862 (1992).
33. J. P. Kumar, D. F. Ready, *Development* **121**, 4359 (1995).
34. C. Gengs *et al.*, *J. Biol. Chem.* **277**, 42113 (2002).
35. C. H. Liu *et al.*, *Neuron* **59**, 778 (2008).
36. J. H. van Hateren, *J. Comp. Physiol. A* **158**, 795 (1986).
37. M. Juusola, G. G. de Polavieja, *J. Gen. Physiol.* **122**, 191 (2003).
38. P. T. Gonzalez-Bellido, T. J. Wardill, M. Juusola, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 4224 (2011).
39. D. A. Clark, L. Bursztyn, M. A. Horowitz, M. J. Schnitzer, T. R. Clandinin, *Neuron* **70**, 1165 (2011).
40. H. Eichner, M. Joesch, B. Schnell, D. F. Reiff, A. Borst, *Neuron* **70**, 1155 (2011).
41. M. V. Srinivasan, *Vision Res.* **25**, 997 (1985).
42. W. Kaiser, in *The Compound Eye and Vision of Insects* (Clarendon, Oxford, 1975), pp. 359–377.

Acknowledgments: We thank G. de Polavieja and people of the Juusola lab for discussions and I. Meinertzhagen and P. Gonzalez Bellido for assistance with electron microscopy. The work was supported by the Gatsby Charitable Foundation (GAT2839: M.J.), the open research fund of the State Key Laboratory of Cognitive Neuroscience and Learning (M.J. and S.T.), the Biotechnology and Biological Sciences Research Council grants (BB/H013849/1 and BB/F012071/1: M.J.) and (BB/G006865/1 and BB/D007585/1: R.C.H.), and the Intramural Research Program of Eunice Kennedy Shriver

National Institute of Child Health and Human Development (HD008776-06: C.H.L.), the NSFC project (30810103906: S.T. and M.J.). The authors declare no competing financial interests. Supplementary materials accompany the paper on *Science* Online. **Author contributions**: Project initiation: M.J.; genetics: T.J.W., O.L., C.H.L., R.C.H., C.J.O'K., X.L.; electrophysiology: X.L., M.J., R.C.H., T.J.W.; behavior: S.D., M.M.C., T.J.W., M.J., S.T.; imaging: O.L., T.J.W., S.D., M.J.; electronmicroscopy: T.J.W.; immunohistochemistry: C.Y.T.; wrote the paper: M.J., with contributions from all authors.

Supplementary Materials
www.sciencemag.org/cgi/content/full/336/6083/925/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S14
Tables S1 and S2
References (*43–105*)

14 October 2011; accepted 13 April 2012
10.1126/science.1215317

AID-Driven Deletion Causes Immunoglobulin Heavy Chain Locus Suicide Recombination in B Cells

Sophie Péron, Brice Laffleur, Nicolas Denis-Lagache, Jeanne Cook-Moreau, Aurélien Tinguely, Laurent Delpy, Yves Denizot, Eric Pinaud, Michel Cogné*

Remodeling of immunoglobulin genes by activation-induced deaminase (AID) is required for affinity maturation and class-switch recombination in mature B lymphocytes. In the immunoglobulin heavy chain locus, these processes are predominantly controlled by the 3′ cis-regulatory region. We now show that this region is transcribed and undergoes AID-mediated mutation and recombination around phylogenetically conserved switchlike DNA repeats. Such recombination, which we term locus suicide recombination, deletes the whole constant region gene cluster and thus stops expression of the immunoglobulin of the B cell surface, which is critical for B cell survival. The frequency of this event is approaching that of class switching and makes it a potential regulator of B cell homeostasis.

Immunoglobulin (Ig) production relies on selection of B cells harboring antigen-specific B cell receptors (BCRs). During antigen-driven responses, Ig genes are reshaped by activation-induced deaminase (AID)–dependent modifications. After somatic hypermutation (SHM) within germinal centers (GCs), B cells with the highest affinity Ig variable (V) domains preferentially capture interactions with T cells. In parallel, AID-dependent class-switch recombination (CSR) diversifies Ig classes. CSR affects the Ig heavy chain (IgH) locus by joining repetitive switch (S) regions that precede Ig isotype–determining constant (C) genes.

Besides the selected winners of random AID-mediated SHM, many cells are obviously losers or undesired responders deserving elimination. Although some unfavorable mutations can promote apoptosis (*1*), abundant survival signals within the GC environment and ligands for nuclear factor κB–activating surface receptors might also activate bystander cells

Limoges University, CNRS, 2 rue Marcland, 87025 Limoges Cedex, France.

*To whom correspondence should be addressed. E-mail: cogne@unilim.fr

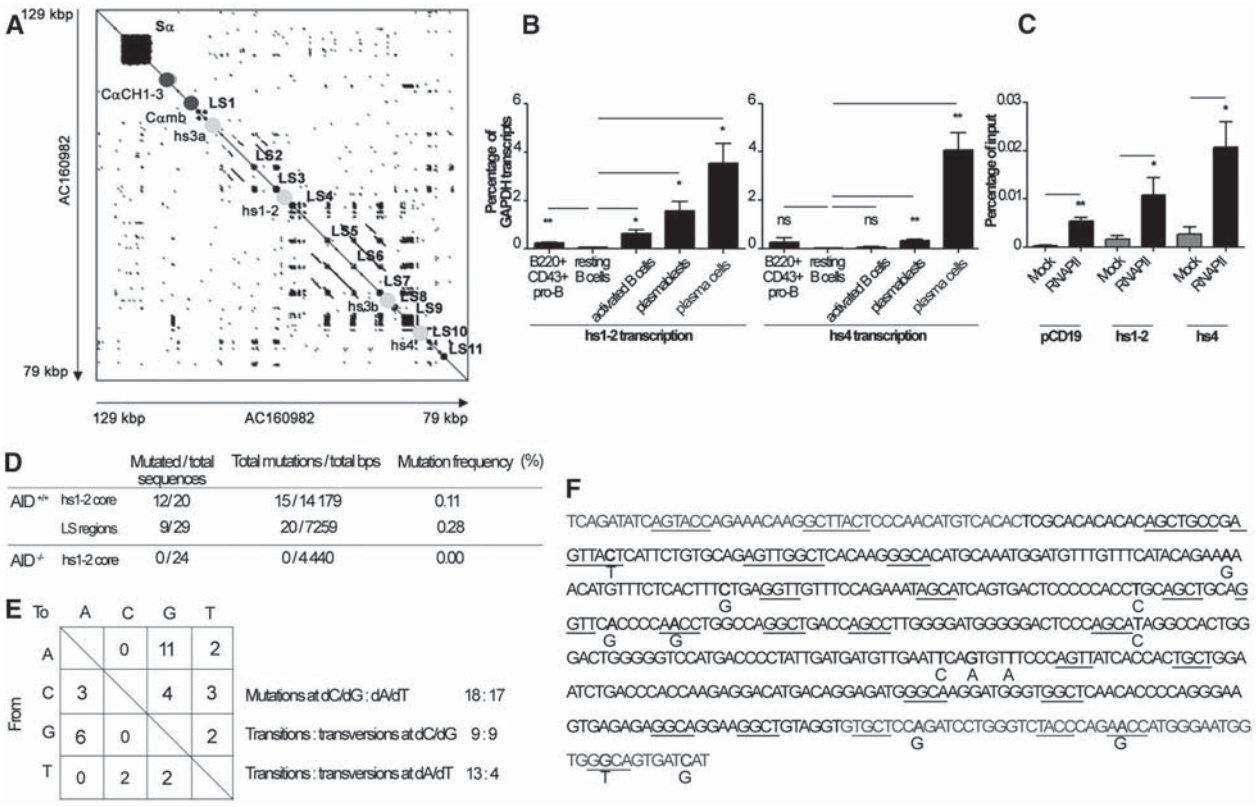


Fig. 1. The IgH 3′RR contains LS sequences targeted by AID. (**A**) LS repeats within the mouse 3′RR. (**B**) Real-time quantitative PCR (qPCR) evaluated hs1-2 and hs4 transcripts in pro-B cells ($n = 2$), resting B cells ($n = 3$), in vitro activated B cells ($n = 7$ for hs1-2 and $n = 3$ for hs4), plasmablasts ($n = 3$), and plasma cells ($n = 5$). (**C**) RNAPII binding. Chromatin immunoprecipitation (ChIP) assay was performed without (mock) or with antibodies against RNAPII in in vitro LPS-stimulated B cells ($n = 3$). Enrichment of hs1-2 and hs4 sequences was assessed by real-time qPCR. CD19 promoter (pCD19) was a positive control. (**D**) Hs1-2 mutation frequency in wild-type (WT) and *Aicda*$^{-/-}$ B cells. (**E**) Hs1-2 mutation spectrum. Number and nature of mutations are depicted. (**F**) Location of mutations. RGYW and WRCY motifs are underlined. For (B) and (C), results are expressed as mean ± SEM. ns, not significant; *$P < 0.05$; **$P < 0.01$ (unpaired two-tailed t test).

with useless BCRs or yield harmful clones that bind self or environmental antigens after random remodeling of the BCR repertoire. Because class-switched antibodies are potent actors of autoimmunity and/or hypersensitivity, means for restricting CSR and reentry of class-switched cells into SHM would help ensure the specificity of immune responses. Yet, how the post-GC repertoire is controlled remains poorly understood.

The mouse IgH locus 3′ regulatory region (3′RR) contains four enhancers (hs3a, 1-2, 3b, and 4) with strong B-lineage specificity (*2*, *3*). It undergoes chromatin remodeling after B cell activation and physical interactions with distant promoters, thus controlling germline transcription and CSR to most C genes and contributing to SHM (*4*–*15*). Both the mouse and human 3′RRs include inverted repeats and multiple stretches of repetitive DNA (*16*–*19*). We analyzed lengths and locations of 3′RR DNA repeats, looking for sequences similar to S regions, which are known to be functionally important (*20*, *21*). In Sγ, 5–base pair (bp) repeats are arranged within higher-order 49-bp repeats. Dot-plot analysis of the mouse 3′RR revealed 11 0.5- to 1-kb-long "like-Sγ" stretches (LS regions) where 49-bp repeats appeared with a threshold set above 60% identity (Fig. 1A). LS1 precedes hs3a; six stretches (LS2 to LS7) are included in the palindrome flanking hs1-2, two blocks (LS8 and LS9) follow hs3b, and another two (LS10 and LS11) follow hs4. We performed similar analyses on human, rabbit, and dog, for which complete IgH locus sequences are available. In all four species, LS regions flank both sides of the 3′RR and are interspersed with enhancers. Human LS repeats are better defined, with greater than 80% identity, whereas rabbit and dog LS regions are highly clustered (rabbit LS4 and dog LS3 are longer than Sα) (fig. S1).

Interspecies conservation of LS repeats suggests that they may have a functional role. Besides their structure, a major feature of S regions is germline transcription before CSR. We thus evaluated mouse 3′RR primary transcription around hs1-2 and hs4 (Fig. 1B). Transcripts were barely detectable in resting B cells but readily appeared upon activation with lipopolysaccharide (LPS). In parallel, chromatin immunoprecipitation (ChIP) experiments confirmed that RNA polymerase II (RNAPII) was associated with 3′ enhancers in activated B cells (Fig. 1C), a finding that was not commented on but recently appeared in a genome-wide analysis of RNAPII load (*22*).

Transcription in activated B cells can result in AID recruitment. V(D)J regions targeted by AID undergo high rates of SHM (above 2% mutations in GC B cells). A much lower off-target rate of AID-mediated mutations affects other non-Ig genes transcribed in B cells (*23*). The 3′ RR elements regulate AID-mediated CSR by physically interacting with AID-targeted IgH promoters (*2*, *6*, *11*, *12*). We wondered whether transcribed 3′RR elements could also be targeted by AID and undergo SHM. In hs1-2 sequences from activated B cells, we found 0.11% SHM (compared with 1.96% for VκJκ SHM in the same samples) (Fig. 1D). Hs1-2 SHM was restricted to AID-sufficient mice and showed no preference for dA/dT or dC/dG residues (Fig. 1E). AID preferentially targets RGYW and WRCY (where R indicates purine; G, guanine; C, cytosine; Y, pyrimidine; W, adenine or thymine) motifs, whereas the base excision repair and mismatch repair processes generate additional mutations. Like V regions, only one-fourth of 3′RR mutations segregated within RGYW or WRCY (Fig. 1F). SHM appeared higher within LS repeats (0.28%) than within core enhancers, approximating the rate observed for Sμ in the same cells (0.42%) (Fig. 1E).

Besides trans-acting factors induced by B cell activation, CSR cis requirements include repetitive DNA, transcriptional accessibility, and AID recruitment (*24*). We thus wondered whether the 3′RR by itself might constitute a substrate for CSR-like recombination and designed primers flanking potential junctions between Sμ and 3′RR LS regions (Fig. 2A). To avoid the multiple tandem or inverted repeats within the 3′RR and to prevent polymerase chain reaction (PCR) artifacts, we focused on two unique regions following the well-defined LS4 and LS10-LS11 stretches (Fig. 2B and table S1). We were able to detect junctions readily and specifically in activated B cells but not in AID-deficient B cells nor in B cells lacking Sμ [α1KI mice (*25*)].

Sμ-to-LS recombination breakpoints were cloned and sequenced (Fig. 3 and fig. S2). By linking the 3′RR and Sμ, the complete C gene

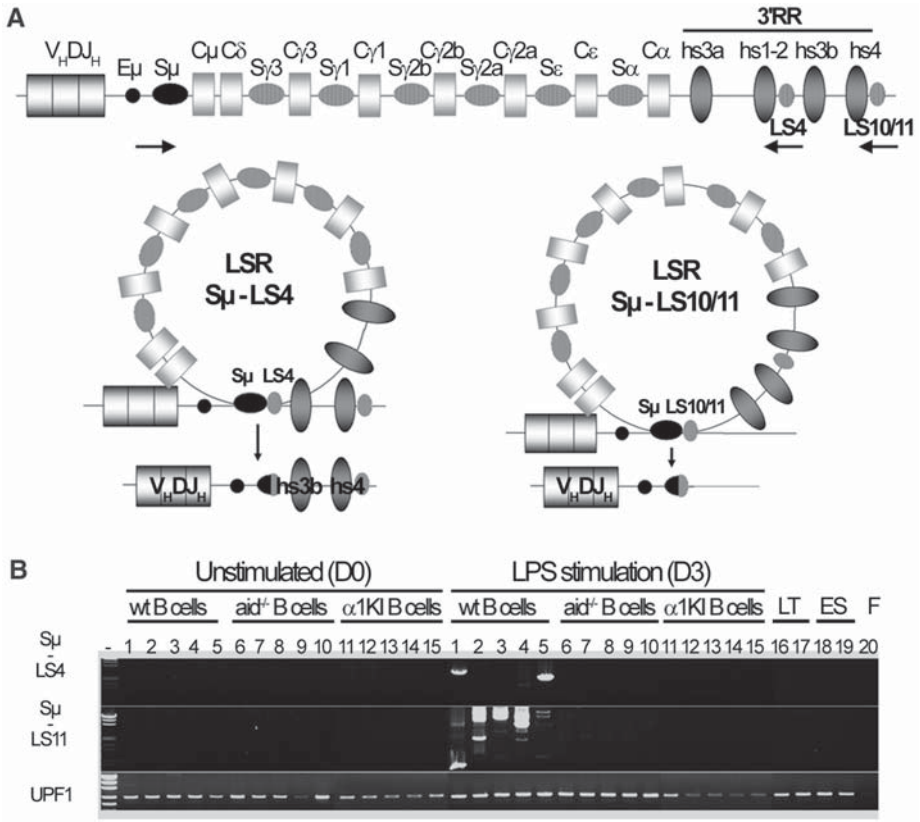


Fig. 2. LSR induction in B cells activated in vitro is AID-dependent. (**A**) LSR between Sμ and LS4 or LS10-LS11. (**B**) LSR was studied by using nested PCR on splenocytes before and after 3 days of LPS activation from the indicated strains of mice (the experiment was repeated 20 times on WT cells, 7 times on AID$^{-/-}$ cells, and 5 times on α1KI cells). α1KI B cells have a deletion of Sμ. UPF1 amplification is a DNA quality control. D0, day 0; D3, day 3; LT, T lymphocytes; ES, embryonic stem cells; F, fibroblasts.

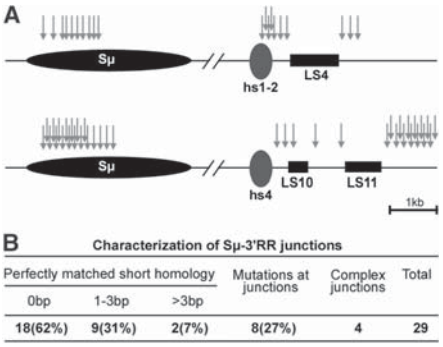


Fig. 3. Characterization of LSR junctions. (**A**) Locations of LSR junction breakpoints (arrows) on Sμ and 3′RR sequences. (**B**) Characterization of Sμ-3′RR junctions.

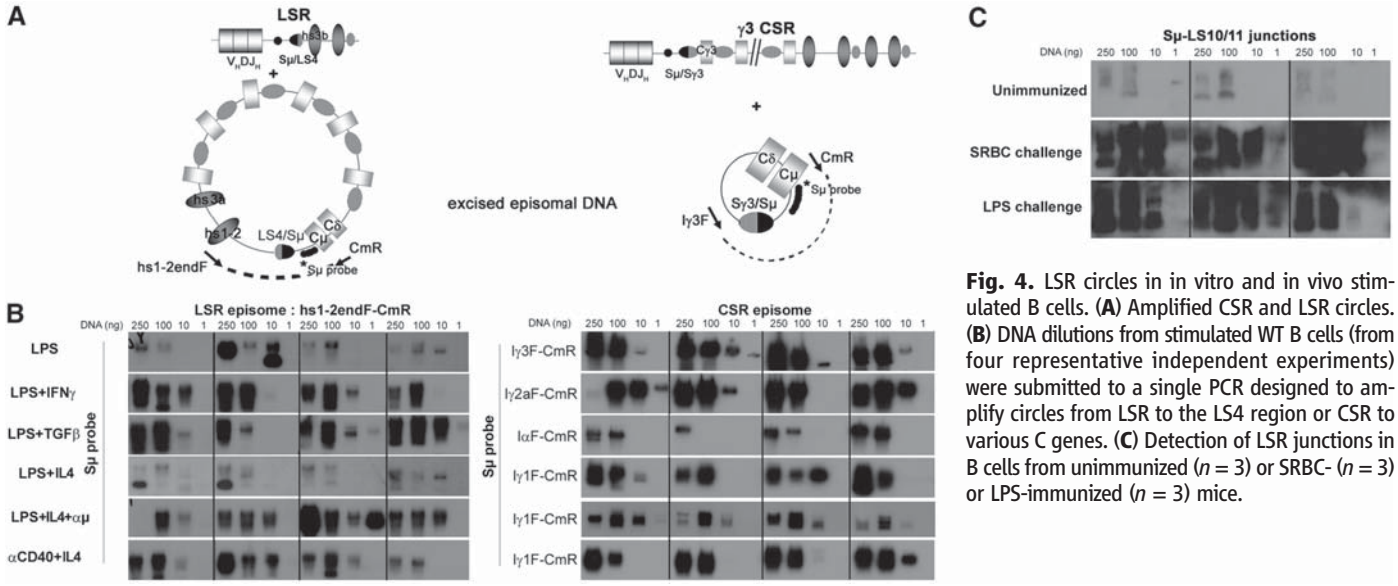


Fig. 4. LSR circles in in vitro and in vivo stimulated B cells. (**A**) Amplified CSR and LSR circles. (**B**) DNA dilutions from stimulated WT B cells (from four representative independent experiments) were submitted to a single PCR designed to amplify circles from LSR to the LS4 region or CSR to various C genes. (**C**) Detection of LSR junctions in B cells from unimmunized ($n = 3$) or SRBC- ($n = 3$) or LPS-immunized ($n = 3$) mice.

cluster is deleted, thereby eliminating BCR expression (or Ig secretion for cells engaged in plasma cell differentiation). Given the importance of BCR expression for B cell survival, we call this process locus suicide recombination (LSR). As for CSR, short junctional microhomologies were occasionally observed. Several complex junctions were characterized: In 2 out of 20 Sμ-LS10-LS11 junctions, an intra-Sμ deletion occurred, whereas another case featured a junction from Sμ to Sγ2b and then from Sγ2b to LS10-LS11. Among 13 Sμ-LS4 junctions sequenced, 1 showed a sequential recombination from Sμ to Sγ3 and then to LS4 (fig. S2).

Because LSR deletes the C gene cluster and is thus expected to compromise B cell survival, comparing frequencies of past lethal LSR versus viable CSR events likely underestimates LSR. Instead, we quantified ongoing recombination by amplifying the circles excised from the IgH locus as by-products of AID-mediated chromosomal deletions (Fig. 4). Such circles are devoid of replication origin, and CSR circles are thus not expected to be amplified with proliferation of class-switched cells. To evaluate LSR frequency, we comparatively amplified CSR circles (joining Cμ to a downstream I promoter) and LSR circles (joining Cμ to hs1-2), in serial dilutions of B cell DNA. In activated B cells, we were able to systematically detect LSR circles in 10 ng of DNA (~2000 cells) or occasionally in 1 ng (~200 cells). By using this approach, we found that, in DNA samples from B cells stimulated under a variety of conditions, LSR circles that contained LS4 were detectable in amounts similar to CSR circles (Fig. 4B). Moreover, LSR is likely underestimated because multiple LS stretches are dispersed throughout the mouse 3′RR and not only downstream of hs1-2. Although we did not find LSR junctions in several commonly used cell lines, including mature B cell and plasma cell lines, they were readily detected in three out of eight IgG class-switched hybridomas, showing that in some cases LSR can only occur on the nonfunctional IgH allele (fig. S3A).

Besides comparing various in vitro B cell stimulation conditions on splenocytes (Fig. 4B), we assayed in vivo stimulated B cells after mouse immunization with LPS or sheep red blood cells (SRBCs) (Fig. 4C). LSR was found in all conditions of in vivo or in vitro B cell activation, and neither cytokines nor BCR cross-linking changed its in vitro occurrence. It was also detected both in marginal zone and follicular B cells sorted from spleens of immunized mice (fig. S3B). Lastly, it was possible by cell cytometry to show one of the likely consequences expected from in vivo occurring LSR, that is, the presence in the spleen of immunized animals of a small proportion of B cells losing BCR expression upon activation (fig. S4). These BCR-negative cells were specifically engaged into apoptosis and significantly differed from BCR-positive cells with regard to annexin V staining (fig. S4B).

AID and the IgH 3′RR are major players in B cell maturation by controlling germline transcription, SHM, and CSR and finally yielding high-affinity BCR and Ig of different isotypes (*2*). We show that the 3′RR itself undergoes germline transcription specifically in activated B cells. Transcription of RNAs from other enhancers ("eRNAs") was recently reported and correlated with enhancer activity (*26*). Similarly to C gene germline transcription, 3′RR eRNAs likely promote AID accessibility. Although the IgH locus forms loops through interactions between the 3′RR and S regions before CSR (*12*, *13*), we show that AID also mutates or breaks the 3′RR itself. DNA synapses may thus either join broken S regions for CSR or Sμ and 3′RR, initiating LSR as a competing process. IgH loops are most likely modulated by costimulatory signals and T-dependent cytokines influencing B cell fate (*12*). Although LSR frequency varied in some in vitro experiments, it was neither stimulated nor inhibited when interleukin 4 (IL4), transforming growth factor–β (TGFβ), or interferon γ (IFNγ) were added, nor by BCR ligation, as far as we can judge by our semiquantitative genetic assay.

How the CSR/LSR balance is finely regulated in vivo remains to be determined. Conserved LS repeats could also serve unknown functions other than LSR, whereas LSR could preferentially target nonfunctional alleles. However, the latter hypothesis is unlikely because LSR on the nonfunctional allele would have little functional implications and is unlikely to justify the evolutionary conservation of LS sequences. Inactivation of one allele in GC B cells lacking recombination activating gene (RAG) activity cannot allow BCR editing either, and we favor the hypothesis that LSR eliminates some mature B cells by targeting the functional allele (eventually being biallelic upon parallel accessibility to recombination). Beyond IgH locus accessibility, cellular interactions within the GC are the fine modulators selecting high-affinity class-switched clones. In this context, LSR might contribute to B cell homeostasis and lead to B cell death when stimulation is suboptimal, whereas efficient cytokine-dependent germline transcription of C genes in an optimal microenvironment would promote CSR and survival of selected antigen-specific B cell clones. LSR could also help shape the peripheral B cell repertoire if it is regulated by BCR antigen specificity.

References and Notes

1. C. C. Goodnow, C. G. Vinuesa, K. L. Randall, F. Mackay, R. Brink, *Nat. Immunol.* **11**, 681 (2010).
2. E. Pinaud *et al.*, *Adv. Immunol.* **110**, 27 (2011).

3. W. Kim, S. Kook, D. J. Kim, C. Teodorof, W. K. Song, *J. Biol. Chem.* **279**, 8333 (2004).
4. V. Giambra *et al.*, *Mol. Cell. Biol.* **28**, 6123 (2008).
5. F. E. Garrett *et al.*, *Mol. Cell. Biol.* **25**, 1511 (2005).
6. W. A. Dunnick *et al.*, *J. Exp. Med.* **206**, 2613 (2009).
7. M. Cogné *et al.*, *Cell* **77**, 737 (1994).
8. J. P. Manis *et al.*, *J. Exp. Med.* **188**, 1421 (1998).
9. A. G. Bébin *et al.*, *J. Immunol.* **184**, 3710 (2010).
10. E. Pinaud *et al.*, *Immunity* **15**, 187 (2001).
11. C. Vincent-Fabert *et al.*, *Blood* **116**, 1895 (2010).
12. R. Wuerffel *et al.*, *Immunity* **27**, 711 (2007).
13. Z. Ju *et al.*, *J. Biol. Chem.* **282**, 35169 (2007).
14. H. Duan, H. Xiang, L. Ma, L. M. Boxer, *Oncogene* **27**, 6720 (2008).
15. M. Gostissa *et al.*, *Nature* **462**, 803 (2009).
16. C. Chauveau, M. Cogné, *Nat. Genet.* **14**, 15 (1996).
17. C. Chauveau, E. Pinaud, M. Cogne, *Eur. J. Immunol.* **28**, 3048 (1998).
18. M. A. Sepulveda, F. E. Garrett, A. Price-Whelan, B. K. Birshtein, *Mol. Immunol.* **42**, 605 (2005).
19. E. Pinaud, C. Aupetit, C. Chauveau, M. Cogné, *Eur. J. Immunol.* **27**, 2981 (1997).
20. A. A. Khamlichi *et al.*, *Blood* **103**, 3828 (2004).
21. R. Shinkura *et al.*, *Nat. Immunol.* **4**, 435 (2003).
22. A. Yamane *et al.*, *Nat. Immunol.* **12**, 62 (2011).
23. M. Liu *et al.*, *Nature* **451**, 841 (2008).
24. J. Stavnezer, J. E. Guikema, C. E. Schrader, *Annu. Rev. Immunol.* **26**, 261 (2008).
25. S. Duchez *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 3064 (2010).
26. T. K. Kim *et al.*, *Nature* **465**, 182 (2010).

Acknowledgments: We thank T. Honjo for providing AID$^{-/-}$ mice and F. Lechouane for sorted B cells DNA samples. We are indebted to the cell sorting facility of Limoges University for excellent technical assistance in cell sorting. This work was supported by grants from Association pour la Recherche sur le Cancer, Ligue Nationale contre le Cancer, Cancéropôle Grand Sud-Ouest, Institut National du Cancer, and Région Limousin. The data presented in this paper are tabulated here and in the supplementary materials.

Supplementary Materials
www.sciencemag.org/cgi/content/full/science.1218692/DC1
Materials and Methods
Figs. S1 to S4
Tables S1 and S2
References (*27*–*30*)

4 January 2012; accepted 27 March 2012
Published online 26 April 2012;
10.1126/science.1218692

Quantitative Sequencing of 5-Methylcytosine and 5-Hydroxymethylcytosine at Single-Base Resolution

Michael J. Booth,[1]* Miguel R. Branco,[2,3]* Gabriella Ficz,[2] David Oxley,[4] Felix Krueger,[5] Wolf Reik,[2,3]† Shankar Balasubramanian[1,6,7]†

5-Methylcytosine can be converted to 5-hydroxymethylcytosine (5hmC) in mammalian DNA by the ten-eleven translocation (TET) enzymes. We introduce oxidative bisulfite sequencing (oxBS-Seq), the first method for quantitative mapping of 5hmC in genomic DNA at single-nucleotide resolution. Selective chemical oxidation of 5hmC to 5-formylcytosine (5fC) enables bisulfite conversion of 5fC to uracil. We demonstrate the utility of oxBS-Seq to map and quantify 5hmC at CpG islands (CGIs) in mouse embryonic stem (ES) cells and identify 800 5hmC-containing CGIs that have on average 3.3% hydroxymethylation. High levels of 5hmC were found in CGIs associated with transcriptional regulators and in long interspersed nuclear elements, suggesting that these regions might undergo epigenetic reprogramming in ES cells. Our results open new questions on 5hmC dynamics and sequence-specific targeting by TETs.

5-Methylcytosine (5mC) is an epigenetic DNA mark that plays important roles in gene silencing and genome stability and is found enriched at CpG dinucleotides (*1*). In metazoa, 5mC can be oxidized to 5-hydroxymethylcytosine (5hmC) by the ten-eleven translocation (TET) enzyme family (*2*, *3*). 5hmC may be an intermediate in active DNA demethylation but could also constitute an epigenetic mark per se (*4*). Levels of 5hmC in genomic DNA can be quantified with analytical methods (*2*, *5*, *6*) and mapped through the enrichment of 5hmC-containing DNA fragments that are then sequenced (*7*–*13*). Such approaches have relatively poor resolution and give only relative quantitative information. Single-nucleotide sequencing of 5mC has been performed by using bisulfite sequencing (BS-Seq), but this method cannot discriminate 5mC from 5hmC (*14*, *15*). Single-molecule real-time sequencing (SMRT) can detect derivatized 5hmC in genomic DNA (*16*). However, enrichment of 5hmC-containing DNA fragments is required, which causes loss of quantitative information (*16*). Furthermore, SMRT has a relatively high rate of sequencing errors (*17*), and the peak calling of modifications is imprecise (*16*). Protein and solid-state nanopores can resolve 5mC from 5hmC and have the potential to sequence unamplified DNA (*18*, *19*).

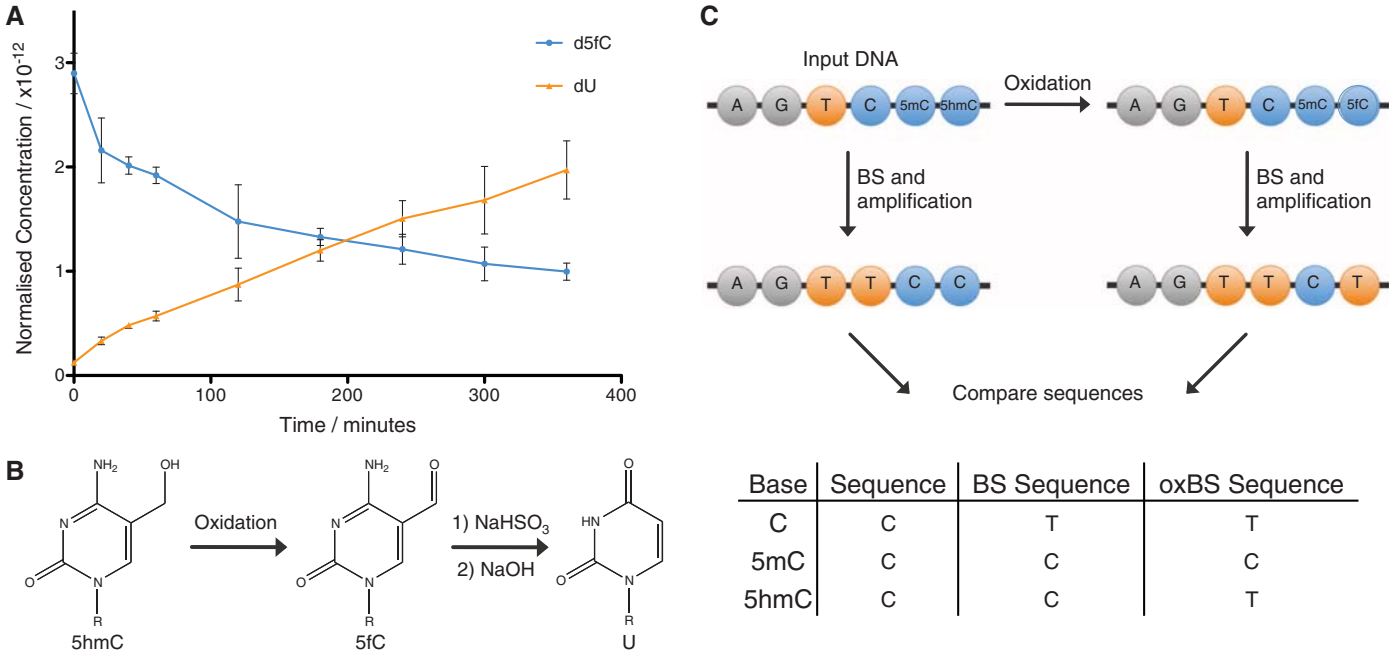
We observed the decarbonylation and deamination of 5-formylcytosine (5fC) to uracil (U) under bisulfite conditions that would leave 5mC unchanged (Fig. 1A and supplementary text). Thus, 5hmC sequencing would be possible if 5hmC could be selectively oxidized to 5fC and then converted to U in a two-step procedure (Fig. 1B). Whereas BS-Seq leads to both 5mC and 5hmC being detected as Cs, this "oxidative bisulfite" sequencing (oxBS-Seq) approach would yield Cs only at 5mC sites and therefore allow us to determine the amount of 5hmC at a particular nucleotide position by subtraction of this readout from a BS-Seq one (Fig. 1C).

Specific oxidation of 5hmC to 5fC (table S1) was achieved with potassium perruthenate ($KRuO_4$). In our reactivity studies on a synthetic 15-nucleotide oligomer single-stranded DNA (ssDNA) containing 5hmC, we established conditions under which $KRuO_4$ reacted specifically with the primary alcohol of 5hmC (Fig. 2A). Fifteen-nucleotide oligomer ssDNA that contained C or 5mC did not show any base-specific reactions with $KRuO_4$ (fig. S1, A and B). For 5hmC in DNA, we only observed the aldehyde (5fC) and not the carboxylic acid (*20*), even with a moderate excess of oxidant. The $KRuO_4$ oxidation can oxidize 5hmC in samples presented as double-stranded DNA (dsDNA), with an initial denaturing step before addition of the oxidant; this results in a quantitative conversion of 5hmC to 5fC (Fig. 2B).

To test the efficiency and selectivity of the oxidative bisulfite method, three synthetic dsDNAs containing either C, 5mC, or 5hmC were each oxidized with $KRuO_4$ and then subjected to a conventional bisulfite conversion protocol. Sanger sequencing revealed that 5mC residues did not convert to U, whereas both C and 5hmC residues did convert to U (fig. S2). Because Sanger sequencing is not quantitative, to gain a more accurate measure of the efficiency of transforming 5hmC to U, Illumina (San Diego, California) sequencing was carried out on the synthetic DNA containing 5hmC (122-nucleotide oligomer) after oxidative bisulfite treatment. An overall 5hmC-to-U conversion level of 94.5% was observed (Fig. 2C and fig. S14). The oxidative bisulfite protocol was also applied to a synthetic dsDNA that contained multiple 5hmC residues (135-nucleotide oligomer) in a range of different contexts that showed a similarly high conversion efficiency (94.7%) of 5hmC to U (Fig. 2C and fig. S14). Last, the $KRuO_4$ oxidation was carried out on genomic DNA and showed through mass spectrometry a quantitative conversion of 5hmC to

[1]Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK. [2]Epigenetics Programme, Babraham Institute, Cambridge CB22 3AT, UK. [3]Centre for Trophoblast Research, University of Cambridge, Cambridge CB2 3EG, UK. [4]Proteomics Research Group, Babraham Institute, Cambridge CB22 3AT, UK. [5]Bioinformatics Group, Babraham Institute, Cambridge CB22 3AT, UK. [6]School of Clinical Medicine, University of Cambridge, Cambridge CB2 0SP, UK. [7]Cancer Research UK, Cambridge Research Institute, Li Ka Shing Centre, Cambridge CB2 0RE, UK.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: wolf.reik@babraham.ac.uk (W.R.); sb10031@cam.ac.uk (S.B.)



Base	Sequence	BS Sequence	oxBS Sequence
C	C	T	T
5mC	C	C	C
5hmC	C	C	T

Fig. 1. A method for single-base resolution sequencing of 5hmC. (**A**) Reaction of 2′-deoxy-5-formylcytidine (d5fC) with $NaHSO_3$ (bisulfite) quenched by NaOH at different time points and then analyzed with high-performance liquid chromatography (HPLC). Data are mean ± SD of three replicates. (**B**) Oxidative bisulfite reaction scheme: oxidation of 5hmC to 5fC followed by bisulfite treatment and NaOH to convert 5fC to U. The R group is DNA. (**C**) Diagram and table outlining the BS-Seq and oxBS-Seq techniques.

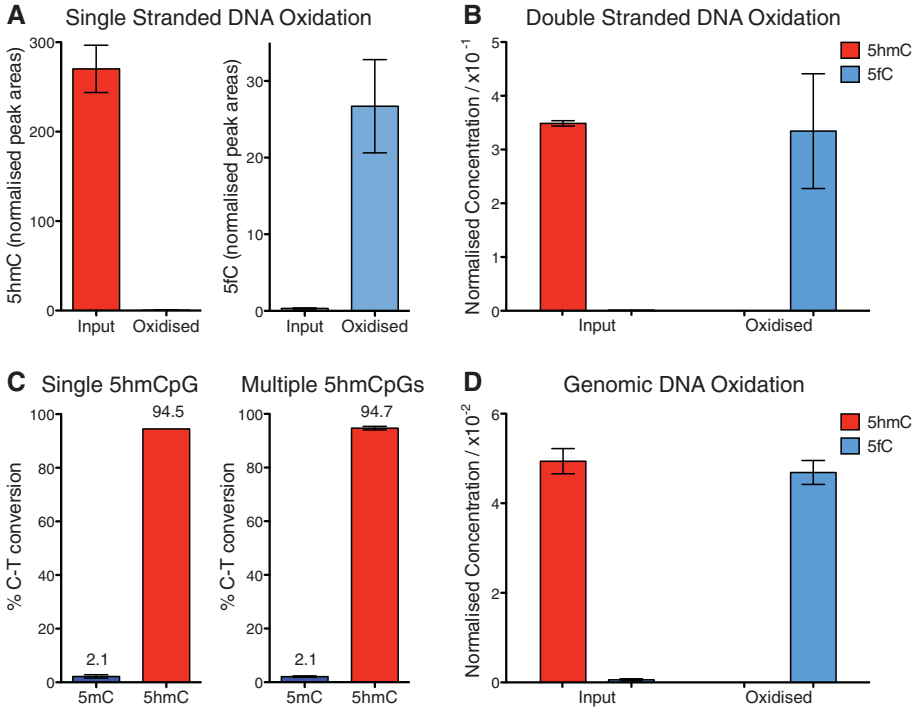


Fig. 2. Quantification of 5hmC oxidation. (**A**) Levels of 5hmC and 5fC (normalized to T) in a 15-nucleotide oligomer ssDNA oligonucleotide before and after $KRuO_4$ oxidation, measured with mass spectrometry. (**B**) Levels of 5hmC and 5fC (normalized to 5mC) in a 135-nucleotide oligomer dsDNA fragment before and after $KRuO_4$ oxidation. (**C**) C-to-T conversion levels as determined by means of Illumina sequencing of two dsDNA fragments containing either a single 5hmCpG (122-nucleotide oligomer) or multiple 5hmCpGs (135-nucleotide oligomer) after oxidative bisulfite treatment. 5mC was also present in these strands. (**D**) Levels of 5hmC and 5fC (normalized to 5mC in primer sequence) in ES cell DNA measured before and after oxidation. Data are mean ± SD.

5fC (Fig. 2D), with no detectable degradation of C (fig. S1C). Thus, the oxidative bisulfite protocol specifically converts 5hmC to U in DNA, leaving C and 5mC unchanged, enabling quantitative, single-nucleotide-resolution sequencing on widely available platforms.

We then used oxBS-Seq to quantitatively map 5hmC at high resolution in the genomic DNA of mouse embryonic stem (ES) cells. We chose to combine oxidative bisulfite with reduced representation bisulfite sequencing (RRBS) (*21*), which allows deep, selective sequencing of a fraction of the genome that is highly enriched for CpG islands (CGIs). We generated RRBS and oxidative RRBS (oxRRBS) data sets, achieving an average sequencing depth of ~120 reads per CpG, which when pooled yielded an average of ~3300 methylation calls per CGI (fig. S3). After applying depth and breadth cutoffs (supplementary materials, materials and methods), 55% (12,660) of all CGIs (*22*) were covered in our data sets.

To identify 5hmC-containing CGIs, we tested for differences between the RRBS and oxRRBS data sets using stringent criteria, yielding a false discovery rate of 3.7% (supplementary materials, materials and methods). We identified 800 5hmC-containing CGIs, which had an average of 3.3% (range of 0.2 to 18.5%) CpG hydroxymethylation (Fig. 3, A and B). We also identified 4577 5mC-containing CGIs averaging 8.1% CpG methylation (Fig. 3B). We carried out sequencing on an independent biological duplicate sample of the same ES cell line but at a different passage

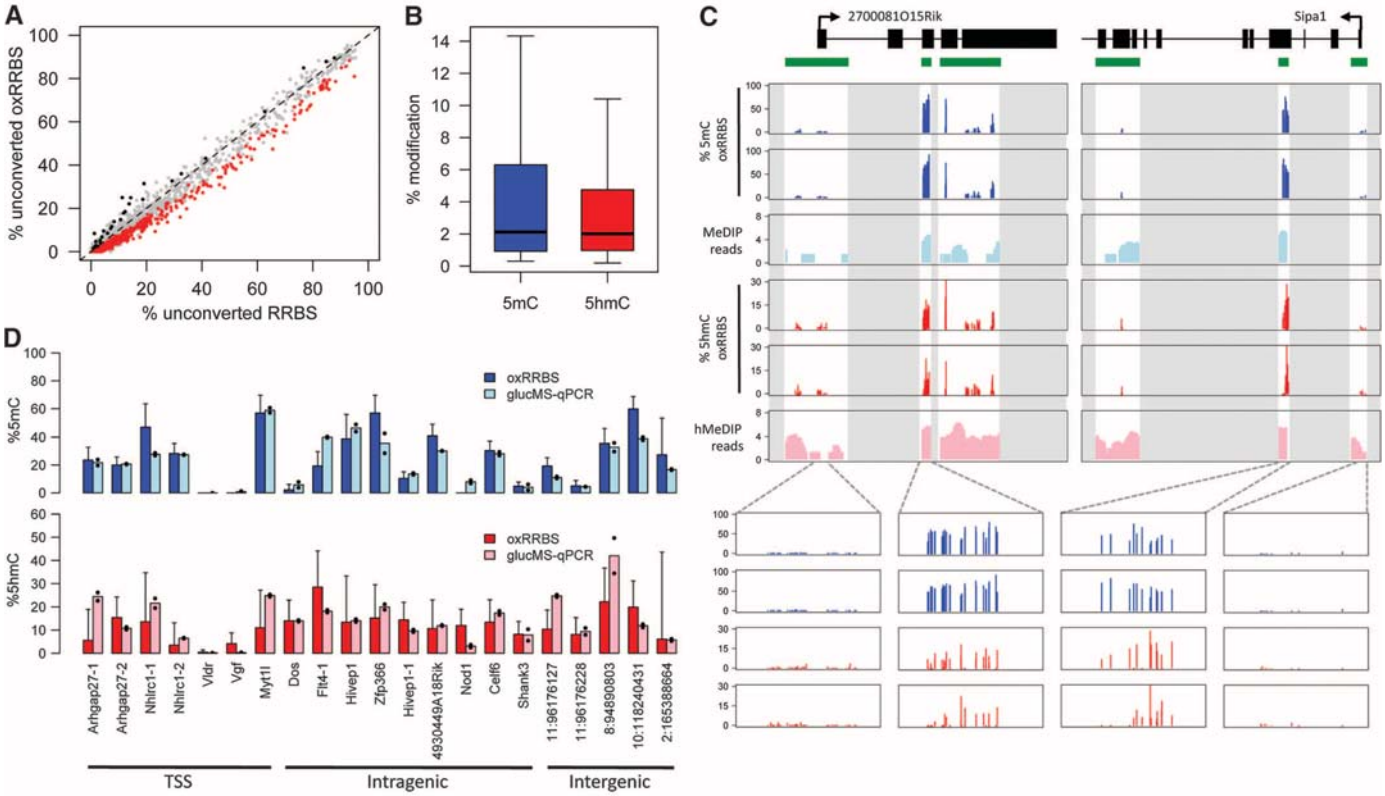


Fig. 3. Quantification of 5mC and 5hmC levels at CGIs by means of oxRRBS. (**A**) Fraction of unconverted cytosines per CGI; 5hmC-containing CGIs (red) have a statistically significant lower fraction in the oxRRBS data set; a false discovery rate of 3.7% was estimated from the CGIs with the opposite pattern (black). (**B**) 5mC and 5hmC levels within CGIs with significant levels of the respective modification. (**C**) Examples of genomic RRBS and oxRRBS profiles overlapped with (h)MeDIP-Seq profiles (*7*). Green bars represent CGIs; data outside CGIs were masked (gray areas). Each bar in the oxRRBS tracks represents a single CpG (in either DNA strand). (**D**) 5mC and 5hmC levels at selected MspI sites were validated through glucMS-qPCR. OxRRBS data are percentage ± 95% confidence interval. Mean glucMS-qPCR values are shown, with the black dots representing individual replicates.

number, which according to mass spectrometry had reduced levels of 5hmC (0.10 versus 0.16% of all Cs), and consistently we found fewer 5hmC-containing CGIs (supplementary text). 5hmC-containing CGIs present in both samples showed good quantitative reproducibility (fig. S5). In non-CpG contexts, we found very few CGIs (71) with levels of 5mC above the bisulfite conversion error (0.2%) (fig. S9) and no CGIs with detectable levels of 5hmC.

Genes associated with 5mC-containing CGIs included *Dazl*, which is known to be methylated in ES cells (fig. S7) (*23*). Similarly, we found that *Zfp64* and *Ecat1* had significant levels of 5hmC (*7*). Genes with >5% 5hmC at transcription start site (TSS) CGIs were associated with gene ontology terms related to transcription factor activity—and in particular were enriched in developmentally relevant genes encoding for Homeobox-containing proteins (such as *Irx4*, *Gbx1*, and *Hoxc4*). To validate our method, we quantified 5hmC and 5mC levels at 21 CGIs containing MspI restriction sites by means of glucosylation-coupled methylation-sensitive quantitative polymerase chain reaction (glucMS-qPCR) (Fig. 3D) (*24*). We found a good correlation between the quantification with oxRRBS and glucMS-qPCR [correlation coefficient (r) = 0.86,

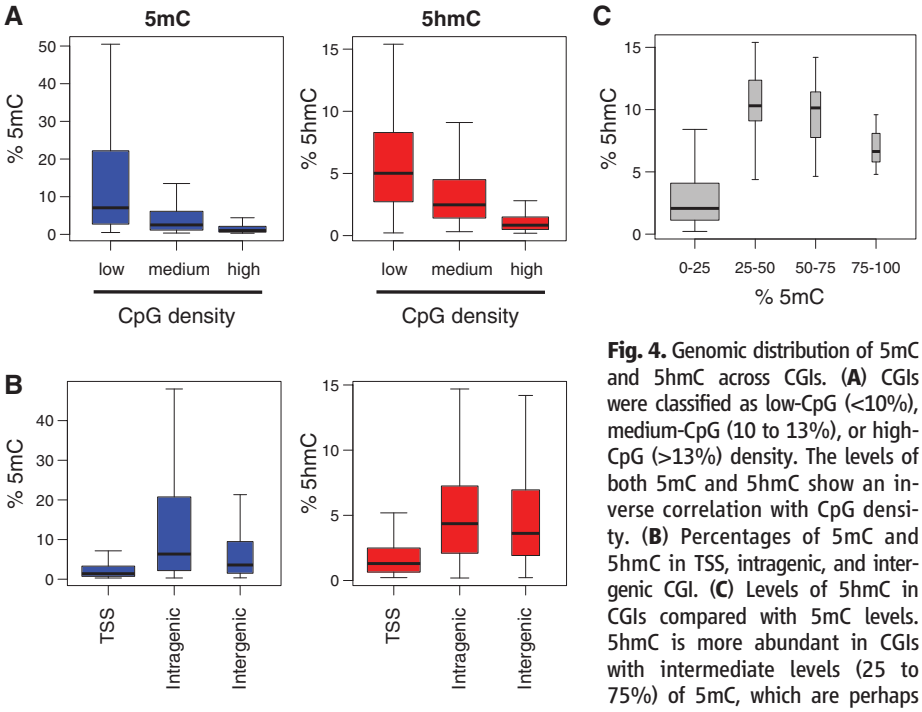


Fig. 4. Genomic distribution of 5mC and 5hmC across CGIs. (**A**) CGIs were classified as low-CpG (<10%), medium-CpG (10 to 13%), or high-CpG (>13%) density. The levels of both 5mC and 5hmC show an inverse correlation with CpG density. (**B**) Percentages of 5mC and 5hmC in TSS, intragenic, and intergenic CGI. (**C**) Levels of 5hmC in CGIs compared with 5mC levels. 5hmC is more abundant in CGIs with intermediate levels (25 to 75%) of 5mC, which are perhaps more epigenetically plastic. For all boxplots, the width of the box is proportional to the amount of data within that group.

$P = 5 \times 10^{-7}$ and $r = 0.52$, $P = 0.01$ for 5mC and 5hmC, respectively], showing that oxRRBS reliably measures 5hmC at individual CpGs. We also found a good correlation between oxRRBS and our previously published (hydroxy)methylated DNA immunoprecipitation sequencing [(h)MeDIP-Seq] data sets (fig. S8) (*7*).

Across CGIs, both 5mC and 5hmC levels are inversely correlated with CpG density, and intragenic and intergenic CGIs contain higher levels of either modification than those overlapping TSSs (Fig. 4, A and B, and fig. S6) (*13*, *22*). TET1 is enriched at TSSs, and thus, a high turnover of 5mC and 5hmC that would keep the steady-state levels low at these sites has been suggested (*9*). Non-TSS CGIs, however, appear to accumulate substantial amounts of both marks, suggesting reduced turnover in these regions. We find that the highest levels of 5hmC are found at CGIs with intermediate levels (25 to 75%) of 5mC (Fig. 4C and fig. S6). Although low-5mC CGIs have reduced potential for 5hmC generation and/or are subjected to a high turnover, high-5mC CGIs are perhaps protected from extensive TET-mediated oxidation, thus stabilizing methylation. Intermediate-5mC CGIs are therefore potentially more epigenetically plastic, given the relatively high abundance of both marks.

Most TSS CGIs (98%) have less than 10% 5mC, as well as low 5hmC, and these are associated with higher transcription levels than average (fig. S10). Within this narrow window, we find a mild negative correlation between transcription and both 5mC and 5hmC levels (fig. S10). At higher 5mC levels, there are insufficient CGIs to obtain a statistically significant result, and it remains possible that here the epigenetic balance between 5mC and 5hmC plays an important transcriptional role, as we previously suggested (*7*).

Last, we quantified 5mC and 5hmC levels at two classes of retrotransposons [long interspersed nuclear element–1 (LINE1) and intracisternal A-particle (IAP)] using two approaches: aligning the oxRRBS reads to the respective consensus sequences and combining oxidative bisulfite with MassARRAY technology (Sequenom, San Diego, California) (fig. S11). We find that LINE1 elements display a considerable amount of 5hmC (approximately 5%), as previously suggested through (h)MeDIP-Seq (*7*). IAPs, on the other hand, have low or no 5hmC. Because LINE1 elements are reprogrammed during preimplantation development whereas IAPs are resistant to this process (*25*), this suggests a possible involvement of 5hmC in the demethylation of specific repeat classes.

The oxBS-Seq method reliably maps and quantifies both 5mC and 5hmC at the single-nucleotide level. Owing to the fundamental mechanism of oxBS-Seq, the approach is compatible with any sequencing platform. In ES cells, we found that in CGIs 5hmC is exclusive to CpG dinucleotides and that it accumulates at intragenic, low-CpG-density CGIs, which tend to have intermediate levels of 5mC and may be particularly epigenetically plastic.

References and Notes

1. A. M. Deaton, A. Bird, *Genes Dev.* **25**, 1010 (2011).
2. M. Tahiliani *et al.*, *Science* **324**, 930 (2009).
3. S. Ito *et al.*, *Nature* **466**, 1129 (2010).
4. M. R. Branco, G. Ficz, W. Reik, *Nat. Rev. Genet.* **13**, 7 (2012).
5. S. Kriaucionis, N. Heintz, *Science* **324**, 929 (2009).
6. M. Münzel *et al.*, *Angew. Chem. Int. Ed.* **49**, 5375 (2010).
7. G. Ficz *et al.*, *Nature* **473**, 398 (2011).
8. W. A. Pastor *et al.*, *Nature* **473**, 394 (2011).
9. H. Wu *et al.*, *Genes Dev.* **25**, 679 (2011).
10. S. G. Jin, X. Wu, A. X. Li, G. P. Pfeifer, *Nucleic Acids Res.* **39**, 5015 (2011).
11. C. X. Song *et al.*, *Nat. Biotechnol.* **29**, 68 (2011).
12. K. Williams *et al.*, *Nature* **473**, 343 (2011).
13. Y. Xu *et al.*, *Mol. Cell* **42**, 451 (2011).
14. Y. Huang *et al.*, *PLoS ONE* **5**, e8888 (2010).
15. C. Nestor, A. Ruzov, R. Meehan, D. Dunican, *Biotechniques* **48**, 317 (2010).
16. C. X. Song *et al.*, *Nat. Methods* **9**, 75 (2012).
17. J. Eid *et al.*, *Science* **323**, 133 (2009).
18. E. V. Wallace *et al.*, *Chem. Commun. (Camb.)* **46**, 8195 (2010).
19. M. Wanunu *et al.*, *J. Am. Chem. Soc.* **133**, 486 (2010).
20. G. Green, W. P. Griffith, D. M. Hollinshead, S. V. Ley, M. Schroder, *J. Chem. Soc. Perkin Trans. 1* **1**, 681 (1984).
21. A. Meissner *et al.*, *Nature* **454**, 766 (2008).
22. R. S. Illingworth *et al.*, *PLoS Genet.* **6**, e1001134 (2010).
23. J. Borgel *et al.*, *Nat. Genet.* **42**, 1093 (2010).
24. S. M. Kinney *et al.*, *J. Biol. Chem.* **286**, 24685 (2011).
25. N. Lane *et al.*, *Genesis* **35**, 88 (2003).

Acknowledgments: We thank T. Green and R. Rodriguez for helpful discussions and J. Webster for help with mass spectrometry. We thank the Biotechnology and Biological Sciences Research Council (BBSRC) for a studentship (M.J.B.). The W.R. lab is supported by BBSRC, Medical Research Council, the Wellcome Trust, European Union EpiGeneSys, and BLUEPRINT. The S.B. lab is supported by core funding from Cancer Research UK. M.J.B. and S.B. are inventors on provisional applications filed for U.S. patents on oxBS-Seq (patent applications US61/605702; US61/641134; US61/623461; and US61/513356). OxRRBS data are deposited in the European Molecular Biology Laboratory–European Bioinformatics Institute ArrayExpress Archive (http://www.ebi.ac.uk/arrayexpress) under the accession number E-MTAB-1042. S.B. is an advisor to Illumina.

Supplementary Materials

www.sciencemag.org/cgi/content/full/science.1220671/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S15
Tables S1 and S2
References (*26*–*40*)

16 February 2012; accepted 13 April 2012
Published online 26 April 2012;
10.1126/science.1220671

New Products

SAMPLE PREPARATION AUTOMATION

SPRIworks HT is a high throughput solution for fragment library preparation on Illumina next generation sequencers. The system allows researchers to process libraries with greater speed and sample reproducibility and decreases processing costs. SPRIworks HT utilizes built-in SPRI (solid phase reversible immobilization)-based per-well size selection. Up to 96 samples can be prepared in as little as three hours without size selection, or six hours with size selection. The simplicity of SPRIworks HT enables users to seamlessly perform automated sample preparation protocols upon their first interaction with the system. SPRIworks HT offers a reagent kit for library preparation that includes sizing solution and polymerase chain reaction reagents. Per-well size selection provides users with the flexibility to process samples for multiple applications in a single run. The overall system increases reproducibility, decreases processing costs, and eliminates most manual preparation steps in next generation sequencing sample prep.

Beckman Coulter

For info: 800-742-2345 | www.beckmancoulter.com

3-D HIGH-CONTENT IMAGING PLATFORM

Designed for demanding high-content assays and cell biology challenges, the new ArrayScan Infinity HCS Reader has 3-D imaging capability and features the latest in variable pin-hole confocal technology, solid-state LED illumination, and live-cell and label-free capabilities as well as best-in-class image analysis and bioinformatics software. The instrument's integrated confocal module features the latest high-speed Nipkow spinning disk technology and a variable pinhole to give finer control over image acquisition across a range of objectives (magnifications), compared to traditional confocal (fixed pin-hole) technology. The confocal module has a four-color LED Light Engine, bringing robust, laser-like performance without the complications and cost of lasers or the need for alignment of camera, confocal, and laser. The ArrayScan Infinity HCS Reader is also coupled with enhanced Thermo Scientific iDev Workflow software for multidimensional analysis. Multiple projection options for image analysis and the ability to save the stack of images to create 3-D movies enable complex 3-D structures to be imaged more clearly.

Thermo Fisher Scientific

For info: 800-432-4091 | www.thermoscientific.com

ELECTRONIC PIPETTE FILLER

The Ultrajet pipette controller is an easy-to-use electronic pipette filler. Ergonomically designed, the Ultrajet is quick and simple to operate and can be used for extended periods of time without discomfort. Large buttons on the hand grip enable precise fine control of aspiration and dispensing of liquids. An integral LCD display shows the aspiration speed which can be simply set with a button under the display. A removable neoprene nozzle grips most standard laboratory pipettes firmly and efficiently allowing users to accurately transfer liquids into plates, vials, or flasks. Compact in size, the Ultrajet is perfect for use in confined spaces such as cabinets and fume hoods. Over filling protection is provided by a disposable 0.45 µm filter in the nose of the controller. Ultrajet uses a rechargeable lithium battery and comes complete with a self-adhesive bench holder for storage.

Porvair Sciences

For info: +44-(0)-1372-824290 | www.porvair-sciences.com

TUMOR CELL PANELS

A collection of Tissue-Specific Tumor Cell Panels combine well-characterized adherent cell lines that were selected for their genomic mutations as found in the Sanger COSMIC database, greatly reducing the time and effort scientists devote to finding these cells. The cell lines are easy to grow using standard media formulations and possess critical genetic abnormalities found in tumors. Each Tissue-Specific Tumor Cell Panel comprises different cell lines that have been grouped by tissue of tumor origin. Descriptive information for each cell line has been annotated with details regarding known mutations in these selected oncogenes or receptors. Cell panels, as research models, are as valuable as the depth of understanding and data supporting them. There are currently 10 Tissue-Specific Tumor Cell Panels in the collection, including panels for breast cancer, triple-negative breast cancer, lung cancer, ovarian cancer, colon cancer, liver cancer, and pancreatic cancer.

ATCC

For info: 800-638-6597 | www.atcc.org/tcp

VISCOSITY BATHS

A new line of viscosity baths are designed to accommodate specific testing and quality control needs, including ASTM D-445. Configured for use with most popular capillary viscometers, these versatile circulating baths feature working temperatures up to 135°C, temperature stabilities to ±0.01°C, and a choice of glass or polycarbonate reservoir. Four different controller types—from an innovative touchscreen model with time/temperature programming capability to an extremely affordable digital temperature controller—are available. All models include a tap-water cooling coil. The Polycarbonate Viscosity Baths feature a 29 L rectangular reservoir with a working depth of 11.25 inches (28.6 cm), square or round viscometer openings, and can be equipped with the Advanced Programmable, Advanced Digital, or Standard Digital Temperature Controller. PolyScience Glass Viscosity Baths are available with either 17 or 27 L round reservoirs and are equipped with the MX Temperature Controller.

PolyScience

For info: 800-229-7569 | www.polyscience.com

Electronically submit your new product description or product literature information! Go to **www.sciencemag.org/products/newproducts.dtl** for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.

cell sciences®

cytokine center

Browse our web site of recombinant proteins, including cytokines, growth factors, chemokines and neurotrophins. Daily shipping and competitive pricing are offered. Bulk quantities of many proteins available. Cell Sciences also carries corresponding antibodies and ELISA kits.

www.cellsciences.com

LIST OF PROTEINS

4-1BBL
4-1BB Receptor
6 Ckine
ACAD8
ACAT2
gAcrp30/Adipolean
Activin A
ACY1
ADAT1
Adiponectin
ADRP
AITRL
Akt1
Alpha-Feto Protein (AFP)
Alpha-Galactosidase A
Angiopoietin-1 (Ang-1)
Angiopoietin-2 (Ang-2)
Angiostatin K1-3
Annexin-V
apo-SAA
Apoliprotein A-1
Apoliprotein E2
Apoliprotein E3
Apoliprotein E4
APRIL
Artemin
ATF2
Aurora A
Aurora B
BAFF
BAFF Receptor
BCA-1 / BLC / CXCL13
BCMA
BD-1
BD-2
BD-3
BDNF
Bivalirudin
BMP-2
BMP-4
BMP-7
BMP-13
sBMPR-1A
Brain Natriuretic Protein
BRAK
Breast Tumor Antigen
C5a
C5L2 Peptide
C-10
C-Reactive Protein
C-Src
Calbindin D-9K
Calbindin D-28K
Calbindin D-29K
Calmodulin
Calcitonin Acetate
Carbonic Anhydrase III
Carcino-embryonic Antigen
Cardiotrophin-1
Caspase-3
Caspase-6
CD4
CD14
CD22
CD40 Ligand / TRAP
CD95 / sFas Ligand
CD105 / Endoglin
CHIPS
CNTF
Collagen
CREB
CTACK/CCL27
CTGF
CTGFL/WISP-2
CTLA-4/Fc
CXCL16
Cytokeratin 8
DEP-1
Desmopressin
Disulfide Oxidoreductase
E-selectin
ECGF
EGF
Elafin/SKALP
EMAP-II
ENA-78
Endostatin
Enteropeptidase
Eotaxin
Eotaxin-2
Eotaxin-3 (TSC)
EPHB2
EPHB4
Eptifbatide
Erk-2
Erythropoietin (EPO)
Exodus-2
Fas Ligand
Fas Receptor
FGF-1 (acidic)
FGF-2 (basic)
FGF-4
FGF-5
FGF-6
FGF-7/ KGF
FGF-8
FGF-9
FGF-10
FGF-16
FGF-17
FGF-18
FGF-19
FGF-20
sFGFR-1 (IIIc) / Fc Chimera
sFGFR-2 (IIIc) / Fc Chimera
sFGFR-3 / Fc Chimera
sFGFR-4 / Fc Chimera
sFlt-1 (native)
sFlt-1 (D3)
sFlt-1 (D4)
sFlt-1 (D5)
sFlt-1 (D7)
Flt3-Ligand
sFlt-4
sFlt-4/ Fc Chimera
Follistatin
FSH
Fractalkine/ CX3C
G-CSF
α-Galactosidase A
Galectin-1
Galectin-3
Gastrointestinal CA
GCP-2
GDF-3
GDF-9
GDF-11
GDNF
GLP-1
Glucagon
Goserelin
GM-CSF
GPBB
GROα
GROβ
GROγ
GRO/MGSA
Growth Hormone
Growth Hormone BP
GST-p21/WAF-1
HB-EGF
HCC-1
HGF
Histdyl-tRNA synthetase
Histrelin
HRG1-β1
I-309
I-TAC
IFN-α
IFN-α A
IFN-α 2a
IFN-α 2b
IFN-β
IFN-γ
IFN-Omega
IGF-I
IGF-II
proIGF-II
IGFBP-1
IGFBP-2
IGFBP-3
IGFBP-4
IGFBP-4
IGFBP-5
IGFBP-6
IGFBP-7
IL-1α
IL-1β
IL-2
IL-3
IL-4
sIL-4 Receptor
IL-5
IL-6
sIL-6 Receptor
IL-7
IL-8 (72 a.a.)
IL-8 (77 a.a.)
IL-9
IL-10
IL-11
IL-12
IL-13
IL-13 analog
IL-15
IL-16 (121 a.a.)
IL-16 (130 a.a.)
IL-17
IL-17B
IL-17D
IL-17E
IL-17F
IL-19
IL-20
IL-22
IL-31
Insulin
IP-10
JE
JNK2a1
JNK2a2
KC / CXCL1
KGF
L-asparaginase
LAG-1
LALF Peptide
LAR-PTP
LC-1
LBP
LD-78β
LDH
LEC/NCC-4
Leptin
LIGHT
LIX
LKM
LL-37
Lymphotactin
sLYVE-1
M-CSF
MCP-1 (MCAF)
MCP-2
MCP-3
MCP-4
MCP-5
MDC (67 a.a.)
MDC (69 a.a.)
MDH
MEC
Mek-1
MIA
Midkine
MIG / CXCL9
MIP-1α / CCL3
MIP-1β / CCL4
MIP-3 / CCL23
MIP-3α / CCL20
MIP-3β / CCL19
MIP-4 (PARC) / CCL18
MIP-5 / CCL15
MMP-3
MMP-7
MMP-13
Myostatin
Nanog
NAP-2
Neurturin
NFAT-1
beta-NGF
NOGGIN
NOV
NP-1
NT-1/BCSF-3
NT-3
NT-4
Ocreotide
Oncostatin M
Osteoprotegerin (OPG)
OTOR
Oxytocin
p38-α
Parathyroid Hormone
PDGF-AA
PDGF-AB
PDGF-BB
Persephin
PF-4
PIGF-1
PIGF-2
PKA α-subunit
PKC-α
PKC-γ
Pleiotrophin
PLGF-1
Polymyxin B (PMB)
PRAS40
PRL-1
PRL-2
PRL-3
Prokineticin-2
Prolactin
Protirelin
PTHrP
PTP1B
PTP-IA2
PTP-MEG2
PTP-PEST
sRANK
sRANKL
RANTES
RELM-α
RELM-β
Resistin
RPTPβ
RPTPγ
RPTPμ
SCF
SCGF-α
SCGF-β
SDF-1α
SDF-1β
Secretin
SF20
SHP-2
STAT1
c-Src
TACI
TARC
TC-PTP
TECK
TFF2
TGF-α
TGF-β1
TGF-β2
TGF-β3
Thymosin α1
sTIE-1/Fc Chimera
sTIE-2/Fc Chimera
TL-1A
TNF-α
TNF-β
sTNFR1
sTNFR2
TPO
TRAIL/Apo2L
sTRAIL R-1 (DR4)
sTRAIL R-2 (DR5)
TSH
TSLP
TWEAK
TWEAK Receptor
Urokinase
VEGF121
VEGF145
VEGF165
VEGF-C
VEGF-C I525
EG-VEGF
VEGF-E
HB-VEGF-E
sVEGFR-1
sVEGFR-2
sVEGFR-3
WISP-1
WISP-2
WISP-3
WNT-1

CELL SCIENCES INC • 480 NEPONSET STREET, BUILDING 12A, CANTON, MA 02021 • INFO@CELLSCIENCES.COM

TOLL FREE: (888) 769-1246 • TEL: (781) 828-0610 • FAX: (781) 828-0542 • WEB: WWW.CELLSCIENCES.COM

NEW ENGLAND BioLabs® Inc.
enabling technologies in the life sciences

Mix it up.

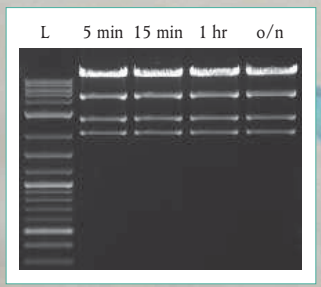
RE-Mix™ Restriction Enzyme Master Mixes

Restriction enzyme digests are now even easier! The same high quality restriction enzymes that you have come to trust from New England Biolabs are now available in master mix format, including enzyme, buffer and loading dye; simply add your DNA and digest.

With RE-Mix Master Mixes take advantage of:

- Simplified and shortened protocols
- Fast digestion in 15 minutes (Time-Saver qualified)
- High product quality with reproducible results

RE-Mix Master Mixes – just add your DNA and mix



pXba DNA was digested with EcoRV-HF™ RE-Mix™ according to the recommended protocol. Lane L is the TriDye™ 2-Log DNA Ladder (NEB #N3270). The same results are obtained whether incubated for 5–15 minutes, 1 hour or overnight.

For more information, visit
www.NEBREMIX.com

GTAGC CTAT
TCUTGA AT G
AGTITC TT
GT TOUT T
AC TACG

To experience the new restriction enzyme challenge from NEB, visit **www.NEBcutitout.com**

Blue-footed Booby (Sula nebouxii)
© Tui De Roy

Enhanced Biological Safety for Flow Cytometry

Innovative solutions that connect with your needs.

Get connected to enhanced biological safety.

Helping all people live healthy lives

Biological safety in flow cytometry is a growing concern for all labs and is another reason you can depend on BD Biosciences to bring together innovative solutions across the spectrum of your needs.

The BioProtect IV from The Baker Company is a biosafety cabinet designed specifically for the BD FACSAria™ product family. Baker has verified that it meets standards for both a Class II Type A2 biosafety cabinet and the National Sanitation Foundation Internal Standard 49.

This enhanced biological safety solution meets a high standard of safety, simplifies cleaning and maintenance, and provides an ergonomic work environment.

Find out how BD Biosciences is addressing biosafety in flow cytometry by delivering interconnected innovations in instrumentation, services, research tools, and partnerships worldwide to help simplify your workflow and protect the integrity of your results.

bdbiosciences.com/go/biosafety

BD Biosciences
2350 Qume Drive
San Jose, CA 95131
bdbiosciences.com

Whoever said ‘It’s the journey not the destination.’

never had to bring a product to market

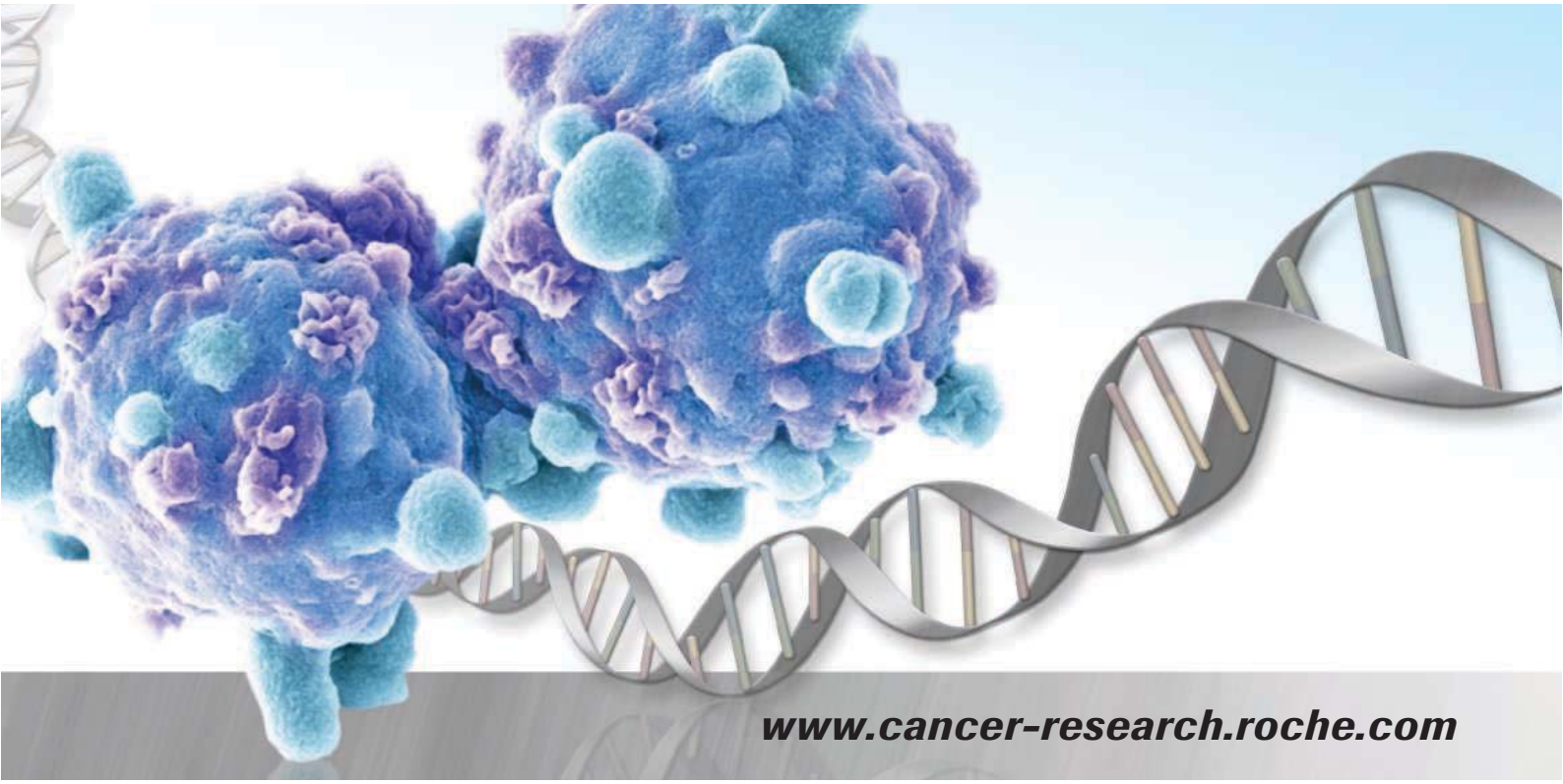
Biopharmaceutical development is difficult. At the core of every successful development program is the ability to produce a sufficient quantity and quality of protein – *rapidly and repeatedly*. Our partners’ ability to rely on the same production platform in discovery, throughout development and into commercialization enables them to focus on developing, *not* producing their product. The unprecedented success rate of *Pf*enex Expression Technology™ makes this a reality.

It’s all about the destination.

pfenex.com

Pfenex Inc.

Your Premier Protein Production Partner



Choose a Trusted, Experienced Partner

To Help You Advance Your Cancer Research Faster

Quantify Gene Expression

Analyze DNA Methylation

Detect and Characterize Genomic Variations

Research Proliferation, Apoptosis, and Cytotoxicity

Continuously Monitor Cell Invasion and Migration

Study the Role of Proteins in Cancer

Hoping to advance your cancer research more quickly? Start by combining **Roche Applied Science**'s world-class reagents and instruments to rapidly, accurately study cancer at the level of the gene, transcript, protein, and cell.

As part of the world's leading supplier of oncology treatments (Roche Pharmaceuticals/Genentech) and tissue-based diagnostics, Roche Applied Science is uniquely qualified to be your primary partner in cancer research, offering:

- **Performance you can trust, plus innovation:** Combine time-tested reagents with novel instrumentation that enables you to study cancer in new ways.
- **Flexible, efficient solutions that help you make more from less:** Obtain more results faster in many applications.
- **Our commitment to you:** Confidently rely on our dedicated service professionals, on-site reagent stocking, customized research solutions, and much more.

Let Roche help you reveal the cellular and molecular mechanisms of cancer. Learn more by visiting *www.cancer-research.roche.com*

Roche Diagnostics Corporation
Roche Applied Science
Indianapolis, Indiana

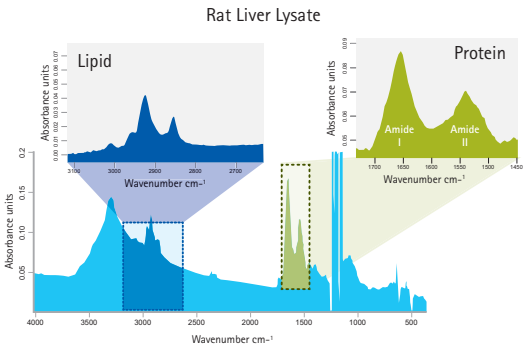
For life science research only.
Not for use in diagnostic procedures.

Goodbye, Bradford Assays!
Drive your research forward with IR-based quantitation.

With Direct Detect™, the first infrared (IR)-based biomolecular quantitation system, there's no sample prep, messy cuvettes or waste—with one-time standard curves. The Direct Detect™ system distinguishes proteins and peptides from interfering sample components, such as lipids. Now you can achieve truly accurate results without the pitfalls of colorimetric assays, even for most lysates and complex samples.

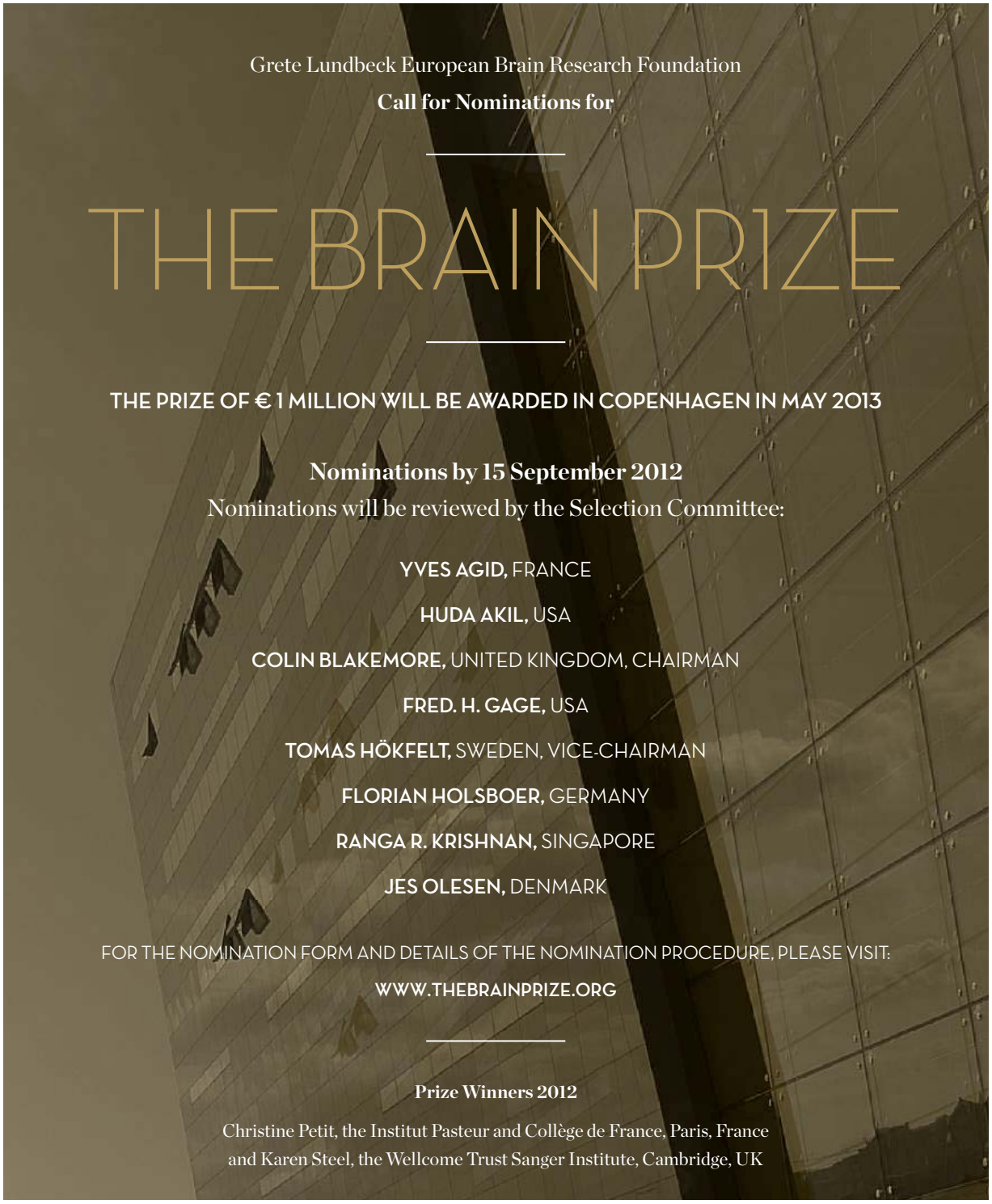
Quit Assays Forever—Quantitate Directly.

Learn more at: **www.millipore.com/DirectDetect**



Accurate IR-based protein quantitation in a lipid-rich lysate. The most intense regions of lipid absorbance are distinct from the protein's Amide I signal.

EMD Millipore is a division of Merck KGaA, Darmstadt, Germany



Grete Lundbeck European Brain Research Foundation

Call for Nominations for

THE BRAIN PRIZE

THE PRIZE OF € 1 MILLION WILL BE AWARDED IN COPENHAGEN IN MAY 2013

Nominations by 15 September 2012

Nominations will be reviewed by the Selection Committee:

YVES AGID, FRANCE

HUDA AKIL, USA

COLIN BLAKEMORE, UNITED KINGDOM, CHAIRMAN

FRED. H. GAGE, USA

TOMAS HÖKFELT, SWEDEN, VICE-CHAIRMAN

FLORIAN HOLSBOER, GERMANY

RANGA R. KRISHNAN, SINGAPORE

JES OLESEN, DENMARK

FOR THE NOMINATION FORM AND DETAILS OF THE NOMINATION PROCEDURE, PLEASE VISIT:

WWW.THEBRAINPRIZE.ORG

Prize Winners 2012

Christine Petit, the Institut Pasteur and Collège de France, Paris, France
and Karen Steel, the Wellcome Trust Sanger Institute, Cambridge, UK

GRETE LUNDBECK
EUROPEAN
BRAIN RESEARCH
FOUNDATION

THE BRAIN PRIZE

The Brain Prize recognizes and rewards outstanding contributions to European neuroscience, from basic to clinical

applied biosystems®
by life technologies™

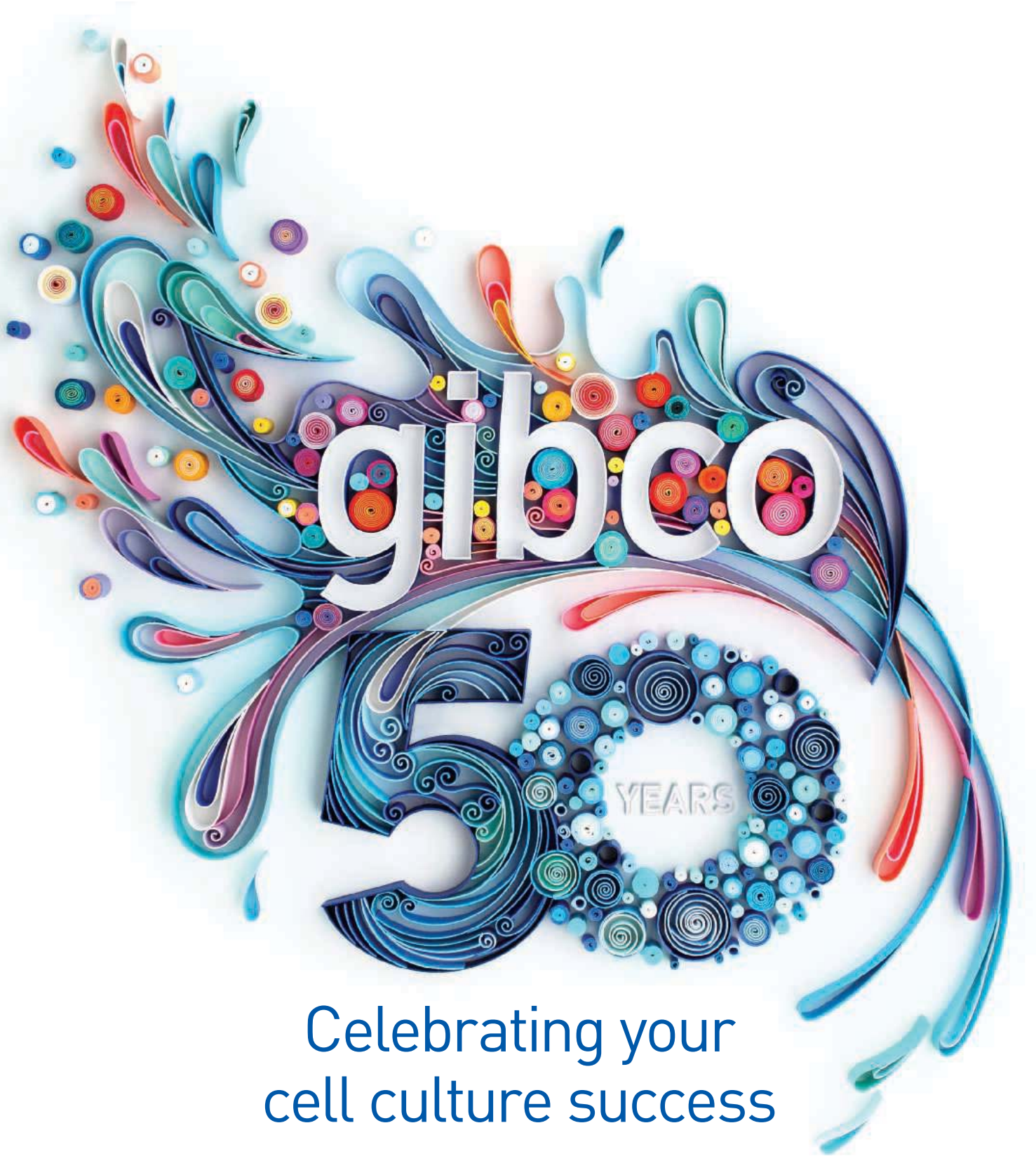
IT FITS

Expanding research on a shrinking budget

Introducing the QuantStudio™ 12K Flex system, the only all-in-one real-time PCR solution designed to scale with your lab. With five interchangeable thermal blocks supporting a variety of TaqMan® and SYBR® Assays, the system allows you to analyze from 1 to over 12,000 data points per run in the format that's right for your project. For ultimate flexibility and gold-standard performance in a single instrument, it's a perfect fit.

Find ultimate flexibility at lifetechnologies.com/quantstudio

life technologies™



Celebrating your cell culture success

In this culture, every little thing matters
lifetechnologies.com/gibco

life technologies™

ion torrent
by life technologies™

PGM™ for genes. Proton™ for genomes.

Sequencing for all.

Powered by fast, simple, scalable semiconductor chips, the Ion PGM™ Sequencer introduced an entirely new approach to sequencing, making it dramatically faster and more accessible.

The new Ion Proton™ Sequencer will go even further. With chip densities up to 1,000 times greater than the Ion PGM™ Sequencer, the Ion Proton™ Sequencer will put whole-genome sequencing within reach of every lab.

Find fast, affordable benchtop sequencing at lifetechnologies.com/ionsequencing

Please visit us at booth 513 at the
Association of Biomolecular Resource Facilities conference.

life technologies™

Be the next winner!

2011 Winner
Dr. Tiago Branco
Postdoctoral
Research Fellow
University College
London

Get recognized!
US$ 25,000 Prize

Deadline for entries:
June 15, 2012

It´s easy to apply! Learn more at:
www.eppendorf.com/prize

eppendorf & Science PRIZE FOR NEUROBIOLOGY

Eppendorf & Science Prize for Neurobiology

Congratulations to Dr. Tiago Branco on winning the 2011 Eppendorf & *Science* Prize for his studies on how dendrites discriminate temporal input sequences and apply different integration rules depending on input location. The results of Dr. Branco's research provide insight on how the brain performs computations, and suggest that even single neurons can solve complex computational tasks.

You could be the 11th winner of this award.
The annual Eppendorf & *Science* Prize for Neurobiology honors young scientists for their outstanding contributions to neurobiological research based on methods of molecular and cell biology. The winner and finalists are selected by a committee of independent scientists, chaired by *Science's* Senior Editor, Dr. Peter Stern.

To be eligible, you must be 35 years of age or younger. If you're selected as this year's winner, you will receive US$ 25,000, have your work published in *Science* and be invited to visit Eppendorf in Hamburg, Germany. Past winners and finalists have come from as far a field as China, Chile, India and New Zealand.

Yes, it can happen to you. Enter your research now!

eppendorf
In touch with life

Science
AAAS

Kurstin Roe Photography

How networked can you get? At Northeastern, very.

Northeastern is the world leader in the surging field of network science. Meet the top scholars driving that success.

Alessandro Vespignani
Sternberg Family Distinguished University Professor of Physics, Computer Science, and Health Science

The recent inductee to Europe's Academy of Science applies his interdisciplinary perspective to forecasting disease epidemics.

David Lazer
Associate Professor of Computer and Information Science and Political Science

A leader in computational social science, Lazer looks at how Internet memes, rumors, and political revolutions spread on Facebook and Twitter.

Ginestra Bianconi
Assistant Professor of Physics

Bianconi, an expert in the physics of complex networks, develops new inference algorithms to analyze dynamic network data.

Alan Mislove
Assistant Professor of Computer and Information Science

The social-networking guru analyzes widespread tweeting patterns to determine trends in national mood and opinion.

Albert-László Barabási
Distinguished Professor of Physics, Biology, and Computer Science

The network-science pioneer and director of Northeastern's Center for Complex Network Research investigates the dynamics of complex systems, such as the Internet and biological networks.

Northeastern University

northeastern.edu

Make your discovery a reality.

University Hospitals Case Medical Center Harrington Discovery Institute is ready to bring your drug discoveries to life, as part of The Harrington Project for Discovery & Development – a $250 million initiative to support the acceleration of medical breakthroughs.

We are proud to introduce the Harrington Scholar-Innovator Grant program, which provides applicants with the opportunity to receive:

- **Grant funding totaling up to $200,000 over two years**
- **Expert mentorship and support**
- **Commercialization assistance to accelerate bringing your breakthrough to market**

Applications are being accepted through **August 1, 2012.**
Apply online at **UHHDI.org/SCI.**

University Hospitals
Case Medical Center

Cleveland | Ohio

Among the nation's leading academic medical centers, University Hospitals Case Medical Center is the primary affiliate of Case Western Reserve University School of Medicine, a nationally recognized leader in medical research and education.

Little Genius

BLItz

BLItz brilliantly packs the power of Dip and Read™ label-free analysis into a personal assay system. Give BLItz a drop of your sample and it does the rest!

- Protein presence/absence in seconds
- Binding kinetics assays at your bench
- Protein quantitation in seconds
- Develop immunoassays in minutes
- Easily analyze crude samples

Cleverly priced under $20K so you can have your own little genius.

Want to try BLItz in your lab? Visit BlitzMeNow.com or call 855.BLITZ.ME.

fortéBIO®
A Division of **Pall Life Sciences**

PALL® Life Sciences

LOCATION: Jackson Park Health Club
ARTICLE: *An Electronic Second Skin*
DATE: Sep 21, 7:43am

LOCATION: University Faculty Lounge
ARTICLE: *The Visual Impact of Gossip*
DATE: Sep 21, 4:22pm

LOCATION: Gyro King
ARTICLE: *Cavemen Craved Carbs, Too*
DATE: Sep 21, 1:13pm

LOCATION: Hemlock Bar
ARTICLE: *Quantum Simulation of Frustrated Classical Magnetism in Triangular Optical Lattices*
DATE: Sep 21, 9:21pm

LOCATION: Bed
ARTICLE: *Consciousness: What, How and Why*
DATE: Sep 21, 10:56pm

A new way to look at science

The new *Science* Reader app for iPad® from AAAS puts *Science* in your hands, wherever you go. Read abstracts, career advice, and highlights from our newest journals, *Science Signaling* and *Science Translational Medicine*. Plus, AAAS members can access full text articles from *Science*. Visit iTunes App Store℠ or content.aaas.org/ipad for details.

Available on the App Store

Science
AAAS

New Products

SAMPLE PREPARATION AUTOMATION

SPRIworks HT is a high throughput solution for fragment library preparation on Illumina next generation sequencers. The system allows researchers to process libraries with greater speed and sample reproducibility and decreases processing costs. SPRIworks HT utilizes built-in SPRI (solid phase reversible immobilization)-based per-well size selection. Up to 96 samples can be prepared in as little as three hours without size selection, or six hours with size selection. The simplicity of SPRIworks HT enables users to seamlessly perform automated sample preparation protocols upon their first interaction with the system. SPRIworks HT offers a reagent kit for library preparation that includes sizing solution and polymerase chain reaction reagents. Per-well size selection provides users with the flexibility to process samples for multiple applications in a single run. The overall system increases reproducibility, decreases processing costs, and eliminates most manual preparation steps in next generation sequencing sample prep.

Beckman Coulter

For info: 800-742-2345 | www.beckmancoulter.com

3-D HIGH-CONTENT IMAGING PLATFORM

Designed for demanding high-content assays and cell biology challenges, the new ArrayScan Infinity HCS Reader has 3-D imaging capability and features the latest in variable pin-hole confocal technology, solid-state LED illumination, and live-cell and label-free capabilities as well as best-in-class image analysis and bioinformatics software. The instrument's integrated confocal module features the latest high-speed Nipkow spinning disk technology and a variable pinhole to give finer control over image acquisition across a range of objectives (magnifications), compared to traditional confocal (fixed pin-hole) technology. The confocal module has a four-color LED Light Engine, bringing robust, laser-like performance without the complications and cost of lasers or the need for alignment of camera, confocal, and laser. The ArrayScan Infinity HCS Reader is also coupled with enhanced Thermo Scientific iDev Workflow software for multidimensional analysis. Multiple projection options for image analysis and the ability to save the stack of images to create 3-D movies enable complex 3-D structures to be imaged more clearly.

Thermo Fisher Scientific

For info: 800-432-4091 | www.thermoscientific.com

ELECTRONIC PIPETTE FILLER

The Ultrajet pipette controller is an easy-to-use electronic pipette filler. Ergonomically designed, the Ultrajet is quick and simple to operate and can be used for extended periods of time without discomfort. Large buttons on the hand grip enable precise fine control of aspiration and dispensing of liquids. An integral LCD display shows the aspiration speed which can be simply set with a button under the display. A removable neoprene nozzle grips most standard laboratory pipettes firmly and efficiently allowing users to accurately transfer liquids into plates, vials, or flasks. Compact in size, the Ultrajet is perfect for use in confined spaces such as cabinets and fume hoods. Over filling protection is provided by a disposable 0.45 µm filter in the nose of the controller. Ultrajet uses a rechargeable lithium battery and comes complete with a self-adhesive bench holder for storage.

Porvair Sciences

For info: +44-(0)-1372-824290 | www.porvair-sciences.com

TUMOR CELL PANELS

A collection of Tissue-Specific Tumor Cell Panels combine well-characterized adherent cell lines that were selected for their genomic mutations as found in the Sanger COSMIC database, greatly reducing the time and effort scientists devote to finding these cells. The cell lines are easy to grow using standard media formulations and possess critical genetic abnormalities found in tumors. Each Tissue-Specific Tumor Cell Panel comprises different cell lines that have been grouped by tissue of tumor origin. Descriptive information for each cell line has been annotated with details regarding known mutations in these selected oncogenes or receptors. Cell panels, as research models, are as valuable as the depth of understanding and data supporting them. There are currently 10 Tissue-Specific Tumor Cell Panels in the collection, including panels for breast cancer, triple-negative breast cancer, lung cancer, ovarian cancer, colon cancer, liver cancer, and pancreatic cancer.

ATCC

For info: 800-638-6597 | www.atcc.org/tcp

VISCOSITY BATHS

A new line of viscosity baths are designed to accommodate specific testing and quality control needs, including ASTM D-445. Configured for use with most popular capillary viscometers, these versatile circulating baths feature working temperatures up to 135°C, temperature stabilities to ±0.01°C, and a choice of glass or polycarbonate reservoir. Four different controller types—from an innovative touchscreen model with time/temperature programming capability to an extremely affordable digital temperature controller—are available. All models include a tap-water cooling coil. The Polycarbonate Viscosity Baths feature a 29 L rectangular reservoir with a working depth of 11.25 inches (28.6 cm), square or round viscometer openings, and can be equipped with the Advanced Programmable, Advanced Digital, or Standard Digital Temperature Controller. PolyScience Glass Viscosity Baths are available with either 17 or 27 L round reservoirs and are equipped with the MX Temperature Controller.

PolyScience

For info: 800-229-7569 | www.polyscience.com

Electronically submit your new product description or product literature information! Go to **www.sciencemag.org/products/newproducts.dtl** for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.

PICTURE YOURSELF AS A AAAS SCIENCE & TECHNOLOGY POLICY FELLOW

Make a Difference.
Help give science a greater voice in Washington, DC! Since 1973, AAAS Fellows have applied their skills to federal decision-making processes that affect people in the U.S. and around the world, while learning first-hand about the government and policymaking.

Join the Network.
Year-long fellowships are available in the U.S. Congress and federal agencies. Applicants must hold a PhD or equivalent doctoral-level degree in any behavioral/social, biological, computational/mathematical, earth, medical/health, or physical science, or any engineering discipline. Individuals with a master's degree in engineering and three years of post-degree professional experience also may apply. Federal employees are not eligible and U.S. citizenship is required.

Apply.
The application deadline for the 2013-2014 AAAS Fellowships is 5 December. Fellowships are awarded in the spring and begin in September. Stipends range from $74,000 to $97,000.

Note: Additional fellowships are available through approximately 30 scientific society partners. Individuals are encouraged to apply with AAAS as well as with any scientific societies for which they qualify.

Full details at: **fellowships.aaas.org**

Enhancing Public Policy, Advancing Science Careers

Sabrina McCormick, PhD

Sociology and Environmental Sciences, Brown University

2009-11 S&T Policy Fellow, Environmental Protection Agency, Office of Research and Development

Now president, Evidence Based Media

AAAS
ADVANCING SCIENCE, SERVING SOCIETY

ELECTROPHORESIS // PROTEAN i12

12 lanes. Individual control. Total confidence.

That's **IEFfective**.

Introducing multilane IEF with unprecedented control. Bio-Rad's PROTEAN® i12™ IEF cell allows you to simultaneously run 12 lanes, each with different samples, protocols, and pH gradients.

Now you can optimize your IEF step in fewer runs or boost lab productivity by performing multiple experiments at once. Set up your run effortlessly using our intuitive touch screen controls.

Independent power supplies allow you to set the precise voltage and current in each lane, so you can walk away with total confidence. Don't let one irregular sample put your entire run at risk.

Take control of your IEF. Call your Bio-Rad sales representative or visit **www.bio-rad.com/ad/IEFfective2** to learn more.

Research. Together.

To find your local sales office, visit **www.bio-rad.com/contact/**
In the U.S., call toll free at 1-800-4BIORAD (1-800-424-6723)

Visit us at **www.bio-rad.com**

BIO-RAD

Tulane University

Post-Doctoral Fellow

A postdoctoral position is available in the Tulane University Department of Pharmacology to conduct NIH-funded research on estrogen receptor signaling, the renin-angiotensin system, and hypertension. The laboratory of Dr. Sarah Lindsey takes an integrative approach by utilizing an in vitro cell culture system, an ex vivo isolated resistance vessel preparation, and in vivo analysis of rodent models. The Post Doc will independently conduct experiments to determine the effect and mechanisms of action of vasoactive drugs on mesenteric arteries. The Post Doc should be familiar with a variety of approaches including isolation of arteries, calcium imaging, histology and immunofluorescence, cell isolation and culture, and Western blotting. Applicants should have a strong background in cardiovascular sciences and preference will be given to individuals with experience handling isolated microvessels and/or culturing vascular smooth muscle or endothelial cells.

Requirements: M.D or D.V.M or Ph.D. or equivalent degree with a strong background in cardiovascular biology research.

Additional Information: Salary and benefits are commensurate with NIH guidelines. Applications should include a brief cover letter highlighting the applicant's research experience and interests, along with full curriculum vitae and the names and contact information of 3 references.

Please submit full curriculum vitae and names of three references via email to: **Sarah Lindsey, Ph.D., Assistant Professor, Department of Pharmacology, Tulane University School of Medicine, 1430 Tulane Avenue, SL 83, New Orleans, LA 70112;** E-mail: **lindsey@tulane.edu. http://tulane.edu/som/departments/pharmacology/lindsey.cfm**. Please also include a letter that summarizes your future research plans. Review of applications will begin immediately and will continue until the position is filled.

Tulane University is an Affirmative Action and Equal Opportunity Employer. We invite Women and Minorities to apply.

Oregon State University OSU

THREE FACULTY POSITIONS IN BIOINFORMATICS AND SYSTEMS BIOLOGY

Oregon State University is conducting three open rank searches for collaboratively-minded faculty who demonstrate enthusiasm for research and teaching at the interface of the quantitative and life sciences, to join our strong growing genomics and systems biology community.

NUCLEIC ACID SEQUENCE BIOINFORMATICS: The appointee will execute a strong research and teaching program at the interface of computer and life sciences that advances the use of nucleic acid sequence data to understand the functions of biological systems at any scale. This is a tenure-track, 9-month, full-time faculty position at any rank that will reside jointly in the School of Electrical Engineering and Computer Science, and in a life science unit depending on the appointee's expertise and interest.

SYSTEMS BIOLOGY OF HOST-MICROBE INTERACTIONS: The appointee will conduct research addressing host-microbe interactions through integrating experimental life science and quantitative methods. The context of the research may be infectious disease, allergic or chronic disease in humans, animals or plants, or mutualistic interactions at the level of individuals or communities. This is a tenure-track, 9-month, full-time faculty position at any rank that will reside jointly in a life science department and in a quantitative science department depending on the appointee's expertise and interest.

APPLIED DISCRETE MATHEMATICS IN SYSTEMS BIOLOGY: The appointee will execute a strong research and teaching program at the interface of applied discrete mathematics and molecular, cellular or organismal systems biology. The context of research may include, but is not limited to, areas of discrete mathematics centered around data integration, network inference, or modeling in systems biology. This is a tenure-track, 9-month, full-time faculty position (Associate or Full Professor) that will reside jointly in the Department of Mathematics, and in a life science unit depending on the appointee's expertise and interest.

More information and application procedures can be found at http://www.cgrb.oregonstate.edu/faculty-positons

GRANTS

みつばち健康科学研究所

5th Yamada Research Grant (Japan)

The Yamada Research Grant aims to promote the latest and highly original studies from multiple points of view in the fields of preventive medicine using bee products, natural products and honeybee science.

1. Eligible researchers:
1) New research grant: For researchers or beekeepers who have never received the research grant from the Yamada Research Grant.
2) Continuing research grant: For researchers or co-researchers who have previously received a research grant from the Yamada Research Grant.

2. Subject for research:
1) New research grant:
(1) Preventive medical research. The following fields will be given a priority.
A. Brain science, neuroscience, mental science
B. Cardiology, gastroenterology, metabolomics (metabolic syndrome, Type 2 diabetes, etc.)
C. Immunology (cancer, allergy, etc.)
D. Otorhinolaryngology, ophthalmology, stomatology
E. Anti-aging medicine (skin, derma, bones, muscles, etc.) or longevity
(2) Basic research on honeybees or beekeeping technology research

2) Continuing research grant:
Progressive research of the theme done with the support of the Yamada Research Grant.

3. Grant amount:
1) New research grant: Up to 2,000,000 Japanese Yen per theme.
2) Continuing research grant: Up to 10,000,000 Japanese Yen per theme.

4. Language: English or Japanese

5. How to apply: Applications should be filled out the specified format on the following website according to application guideline for this grant.

The official website of Application for Yamada Research Grant 2012:
http://grant.bee-lab.jp/grant/grant_2012/guideline_eng.html

6. Application period: From May 1, 2012 to June 29, 2012 (until 17:00 Japan Time)

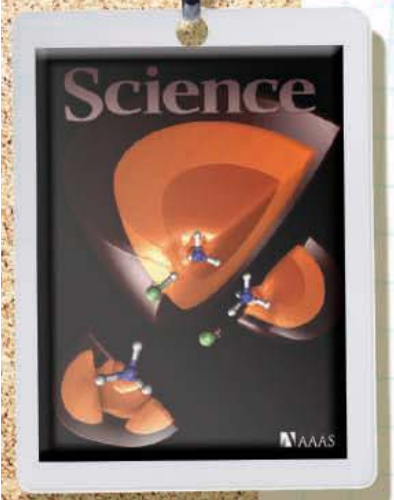
How to inquire: If there are any questions, please contact the secretariat of this grant by e-mail. E-mail address: **bee-lab@yamada-bee.com**

Get a Career Plan that Works.

An exceptional career requires insightful planning and management. That's where *Science* Careers comes in. From job search to career enhancement, *Science* Careers has the tools and resources to help you achieve your goals. Get yourself on the right track today and get a real career plan that works. Visit **ScienceCareers.org**.

Science Careers
From the journal *Science* AAAS

ScienceCareers.org



Find My Job

Develop My Career

Research

Navigate

Network

Search

Track

UPLOAD A RESUME

AAAS is here – helping scientists achieve career success.

Every month, over 400,000 students and scientists visit ScienceCareers.org in search of the information, advice, and opportunities they need to take the next step in their careers.

A complete career resource, free to the public, *Science* Careers offers a suite of tools and services developed specifically for scientists. With hundreds of career development articles, webinars and downloadable booklets filled with practical advice, a community forum providing answers to career questions, and thousands of job listings in academia, government, and industry, *Science* Careers has helped countless individuals prepare themselves for successful careers.

As a AAAS member, your dues help AAAS make this service freely available to the scientific community. If you're not a member, join us. Together we can make a difference.

To learn more, visit **aaas.org/plusyou/sciencecareers**

AAAS + U = Δ

POSITIONS OPEN

PROFESSOR AND DIRECTOR for the Center for Sensorimotor Neural Engineering

The University of Washington invites applications for a full time tenure-track position as Professor within the College of Engineering. We are seeking an outstanding leader who will serve as the Director of the Center for Sensorimotor Neural Engineering (**website: http://www.csne-erc.org**) a National Science Foundation funded Engineering Research Center that is dedicated to a cross-disciplinary focus on technologies that assist individuals with mobility-related neural disorders and devices that interact with, and are inspired by, nervous system function. The Director will have a vision for harnessing the tremendous potential of Neural Engineering for societal impact, medicine, manufacturing, education, and other emerging areas.

The CSNE Director is responsible for fostering the growth of a high impact Neural Engineering program and managing the research, educational, and industry components of a Gen3 ERC. Responsibilities include fundraising from state and federal agencies, industry, and foundations; fiscal management; operational and strategic planning; establishment, implementation, and measurement of goals and objectives; implementation of policy and procedures; and partnership strengthening and communication. The CSNE Director will report to the Dean of the College of Engineering. Potential home departments include Computer Science & Engineering, Bioengineering, Electrical Engineering, and Mechanical Engineering.

Review of applications will begin July 2012, and will be accepted until December 2012.

To apply please visit **website: http://www.washington.edu/admin/acadpers/ads/aa3112.html.**

Please send questions or requests for information to **Dr. Tom Daniel,** Search Committee Chair, at **e-mail: csne@uw.edu.** Thank you for your interest in this position at the University of Washington.

The University of Washington is building a culturally diverse faculty and strongly encourages applications from women and minority candidates. The University is an Affirmative Action/Equal Opportunity Employer. The University of Washington, a recipient of the 2006 Alfred P. Sloan award for Faculty Career Flexibility, is committed to supporting the work-life balance of its faculty. University of Washington faculty engage in teaching, research, and service.

The Center of Mitochondrial Diseases Case Western Reserve University School of Medicine

The Center for Mitochondrial Diseases at Case Western Reserve University invites applications for a non-tenure-track position at the **ASSISTANT PROFESSOR** level.

Applicants must have demonstrated scholarly achievement in research in mitochondrial physiology and mitochondrial bioenergetics. The successful candidate is expected to advance independently in a field of research, maintain an independent funded research program, and be actively involved in the diagnosis and treatment of human mitochondrial diseases. Curriculum vitae, a statement of research interest, and three letters of recommendation should be submitted to **Cami Thompson** (**e-mail: cami.thompson@case.edu**).

In employment, as in education, Case Western Reserve University is committed to Equal Opportunity and Diversity. Women, veterans, members of underrepresented minority groups, and individuals with disabilities are encouraged to apply. Case Western Reserve University provides reasonable accommodations to applicants with disabilities. Applicants requiring a reasonable accommodation for any part of the application and hiring process should contact the Office of Inclusion, Diversity, and Equal Opportunity at ***telephone: 216-368-8877*** *to request a reasonable accommodation. Determinations as to granting reasonable accommodations for any applicant will be made on a case-by-case basis.*

POSITIONS OPEN

ASSISTANT PROFESSOR Systems Physiology The Division of Basic Biomedical Sciences Sanford School of Medicine of The University of South Dakota

The Division of Basic Biomedical Sciences at the University of South Dakota's Sanford School of Medicine of invites applications for a tenure-track faculty position at the Assistant or Associate Professor level. Applicants should have a Ph.D. or M.D.-Ph.D. in Physiology, Pharmacology or other related fields and at least two years of postdoctoral experience. We seek energetic, interactive individuals whose research interests include, but are not limited to, angiogenesis, vascular biology, or pharmacology. Preference will be given to candidates whose research complements existing strengths in the Division. (See **website: http://www.usd.edu/medical-school/biomedical-sciences/faculty-staff.cfm** for a complete list of faculty within the Division of Basic Biomedical Sciences.) Successful candidates will be expected to develop an independent, externally funded research program and to participate in teaching graduate or medical students. Excellent startup funds, state-funded salary commensurate with experience and state of the art research facilities in the new Lee Medical Science Building (Vermillion, South Dakota) will be provided. The application should include curriculum vitae, a summary of past research and teaching experience, a statement of research interests and future plans as well as the names of three references. All materials should be sent to The University of South Dakota online employment **website: https://yourfuture.sdbor.edu.** Review of applications will begin on June 18, 2012 and continue until position is filled. *Women and minorities are encouraged to apply. Affirmative Action/Equal Opportunity Employer.*

ASSISTANT PROFESSOR, TENURE-TRACK Genomics and Bioinformatics

The Department of Biology, The University of Texas at Tyler, invites applications for a tenure-track Assistant Professor position. The incumbent will establish an extramurally funded research program in animal or plant genomics and bioinformatics and teach undergraduate and graduate courses. Qualifications: Ph.D. in genomics and bioinformatics, postdoctoral experience, strong publication record, grantsmanship, and teaching experience. Applications must include (as one PDF file) curriculum vitae, statements of research interests and teaching philosophy, reprints of three publications and sent to **Dr. Lance Williams,** Chair, Genomics Search at **e-mail: lwilliams@uttyler.edu.** Three reference letters must be sent to the same address. Screening will begin July 31, 2012 and continue until a suitable candidate is identified. This is a security sensitive position and subject to criminal history check. Full job description **website: http://www.uttyler.edu/biology.** *The University of Texas is an Equal Opportunity Employer; women and minorities are encouraged to apply.*